APTD-0788

PB 204 714

COMPUTER MAPPING
as an aid in
AIR POLLUTION STUDIES

VOLUME TWO: INDIVIDUAL REPORTS

REPORT M
COMPUTER EQUIPMENT AND PROGRAMS

by

Allan H. Schmidt

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
Springfield, Va. 22151

A

84

BIBLIOGRAPHIC DATA SHEET	1. Report No. APTD-0788	2.	3. Recipient's Accession No.
4. Title and Subtitle Computer Mapping as an aid in Air Pollution Studies Volume Two: Individual Reports - Report M - Computer Equipment and Programs			5. Report Date April 1970 6.
7. Author(s) Allan H. Schmidt			8. Performing Organization Rept. No.
9. Performing Organization Name and Address Graduate School of Design Harvard University Cambridge, Massachusetts 02138			10. Project/Task/Work Unit No. 11. Contract/Grant No. 68A-2405D
12. Sponsoring Organization Name and Address National Air Pollution Control Administration U. S. Public Health Service U. S. Department of Health, Education, and Welfare Washington, D. C. 20201			13. Type of Report & Period Covered 14.

15. Supplementary Notes - DISCLAIMER - This report was furnished to the Office of Air Programs by Graduate School of Design, Harvard University, Cambridge, Massachusetts 02138 in fulfillment of Grant No. 68A-2405D

16. Abstracts

The report presents an evaluation of computer equipment and programs for graphic display. It is divided into three sections; computing machinery (hardware), computer programs (software) and types of maps, and data for graphic display. In addition, the introduction discusses the criteria for computer mapping, principles of operation, and current and future applications.

17. Key Words and Document Analysis. 17a. Descriptors

Air pollution
Computer programs
Mapping
Computer systems (hardware)
Computer systems programs
Maps

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group 13B

18. Availability Statement Unlimited	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 86
	20. Security Class (This Page UNCLASSIFIED	22. Price

FORM NTIS-35 (10-70) USCOMM-DC 40329-P71

LABORATORY FOR COMPUTER GRAPHICS
AND SPATIAL ANALYSIS

Graduate School of Design
Harvard University
Cambridge, Massachusetts

This report was funded by:

Demonstration Grant 68A-2405D
National Air Pollution Control Administration
U. S. Public Health Service
U. S. Department of Health, Education, and Welfare

John C. Goodrich, Editor
Gail Howrigan, Associate Editor

Rosalie Alexander
Carol Chase
Candace Howes
Whitmore John
Linda Peacock

April 1970

COMPUTER MAPPING

as an aid in

AIR POLLUTION STUDIES

Personnel

Principal Investigators

Howard T. Fisher	Project Director (Associate Director of The Laboratory for Computer Graphics and Spatial Analysis)
John C. Goodrich	Associate Project Director
William Warntz	(Director of The Laboratory for Computer Graphics and Spatial Analysis)
Allan H. Schmidt	(Associate Director of The Laboratory for Computer Graphics and Spatial Analysis)

Case Study Investigators

Norma Bogen	St. Louis Project
Paul Jack Dangermond	California and Puget Sound Studies
Steven Fosburg	Statistical Investigation
Jean Maurice Granger	Montreal Project
Donald Shepard	St. Louis Project/Program Investigation
Elia Werczberger	Kansas City Project

Programming Support and Development

Chang-I Hua
Lannon Leiman
Kathleen Reine

Research Assistants

Whitmore John
Johnny Kirkland
Helen Mansfield
Adrian Thomas

Ellen Bussard
Robert Cartwright
Katherine Mickey
Warnie Webster

Administrative and Secretarial Support

Malcolm Fitzpatrick
Gail Howrigan
Linda Peacock
Laurence Yont

Regina McDonald
Donna L. Moorhead
Claudia S. Rosen
Judith Strong

PREFACE

From July 1, 1968 through December 31, 1969, the Laboratory for Computer Graphics and Spatial Analysis conducted a demonstration of several conceptual, methodological, and technical approaches to problems in the field of air pollution, employing computer-produced maps and allied displays. This work was done under a grant (No. 68A-2405D) from the Division of Control Agency Development, National Air Pollution Control Administration (NAPCA), Public Health Service, U. S. Department of Health, Education, and Welfare.

The results of this study, taken in its entirety, should make an important contribution to the work of Federal, state, and local air pollution agencies, especially in their efforts to determine the best control strategies for use in air quality control regions. The computer is an ideal tool for relating and combining sets of data, and for considering alternatives. For instance, graphic displays can contribute significantly to the evaluation of both short-term and long-term control strategies.

In demonstrating "the use of computer graphics as a means for more clearly portraying data related to the sources, distribution, and community effects of air pollution," the primary emphasis was on case studies for specific areas. The secondary emphasis, to support and complement the case studies, included:

(a) "the development of instructional manuals covering the procedures required for the preparation and interpretation of computer maps of various kinds, as related to air pollution,"

(b) "the use of certain newly developed techniques for describing, analyzing, and portraying quantitative surfaces and movements relative to such surfaces, as related to air pollution," and

(c) "an appraisal of alternative equipment, techniques, and computer programs presently or potentially available for the production of computer graphic displays pertinent to air pollution."

The case study regions were selected in consultation with NAPCA. Considerations included:

(a) sufficient air quality, emission, and related socio-economic and meteorological data--spatially distributed over the region--to enhance the likelihood of adequately demonstrating the mapping techniques;

(b) representative problems in data handling and mapping so that the techniques demonstrated would have applicability to as many other areas as possible;

(c) special problems in data handling and mapping so that as many techniques as possible could be adapted, developed, and tested;

(d) areas of which NAPCA and the individual investigators had special knowledge and could therefore utilize relevant research to enhance the study;

(e) areas where an immediate interest in mapping had been expressed for on-going Federal, state, and local air pollution activities.

Other research was undertaken to support and complement the case studies, and, in particular, to provide:

(a) the widest possible distribution of the mapping programs and accompanying explanation of the techniques through

i) copies of the Laboratory's SYMAP Computer mapping program,
ii) an introductory correspondence course in the SYMAP program,
iii) general program description and supplementary explanation for the existing user-manuals;

(b) a general review of present and potential hardware, output devices, and programs applicable to computer graphic displays of air pollution and related data;

(c) means for predicting, projecting, and correlating air pollution and related data, utilizing the latest

knowledge in planning, geography, and spatial analysis in general;

(d) dissemination of information on approaches and techniques to Federal, state, and local air pollution agencies, especially for use in the air quality control regions.

The final report is divided into two parts: Volume One: Summary, and Volume Two: Individual Reports--the detailed reports on the case study areas and the technical reports related to other research undertaken. The pertinent graphic displays are included with the appropriate individual report and a complete set is provided with the Summary Volume.

Introductory sections describing the air pollution and related data used are provided in the major case study reports for St. Louis and Kansas City; these should be referred to by anyone unfamiliar with the types of data and problems existing in this field. Detailed discussions of special data handling and mapping techniques are presented either as appendices to case study reports, or in separate technical reports.

The Summary Volume includes a discussion of all the areas of research undertaken and contains the complete table of contents for the reports of Volume Two. This should provide a useful reference of the special problems and technical discussions of one report which may be of interest to the reader of another detailed report.

The evaluation of computer equipment and programs for graphic display is presented here as Report M of Volume Two. It is divided into three sections--computing machinery (hardware), computer programs (software) and types of maps, and data for graphic display--in addition to an introduction discussing the criteria for computer mapping, principles of operation, and current and future applications.

The report was prepared by Allan H. Schmidt, with the assistance of Adrian Thomas and John Goodrich.

John C. Goodrich, Editor
Cambridge, Massachusetts
April, 1970

REPORT M

COMPUTER EQUIPMENT AND PROGRAMS

Table of Contents

PREFACE

1. INTRODUCTION

1.1. Criteria for Computer Mapping
1.1.1. Map Characteristics
1.1.2. Data Characteristics

1.2. Principles of Operation
1.2.1. Input
1.2.2. Processing
1.2.3. Output

1.3. Current and Future Applications
1.3.1. Military and Aerospace Applications
1.3.2. Costs Involved
1.3.3. Other Current Applications
1.3.4. Future Applications

1.4. Air Pollution Considerations

2. COMPUTING MACHINERY (HARDWARE)

2.1. Computers

2.2. Input Devices

2.3. Output Devices
2.3.1. Printers
2.3.2. Mechanical Plotters
2.3.3. Cathode Ray Tubes

3. COMPUTER PROGRAMS (SOFTWARE) AND TYPES OF MAPS

3.1. Types of Maps
3.1.1. Contour Maps
3.1.2. Conformant Maps
3.1.3. Dot Maps
3.1.4. Symbol Maps
3.1.5. Three-Dimensional Maps
3.1.6. Four-Dimensional Maps
3.1.7. Color
3.1.8. Non-Geographic Mapping

3.2. Computer Programs
3.2.1. Plotter Programs

3.2.2. Printer Programs
3.2.3. SYMAP
3.2.4. SYMVU
3.2.5. Other Mapping Programs

3.3. Air Pollution Considerations
3.3.1. Types of Maps
3.3.2. Computer Programs

4. DATA FOR GRAPHIC DISPLAY

4.1. Types of Data
4.1.1. Point or Node Data
4.1.2. Line or Segment Data
4.1.3. Area or Zone Data
4.1.4. Incompatible Data Zones

4.2. Data Storage
4.2.1. For Point Data
4.2.2. For Line Data
4.2.3. For Area Data
4.2.4. Conclusions

4.3. Air Pollution Considerations
4.3.1. Types of Data
4.3.2. Data Storage
4.3.3. Data Manipulation

REFERENCES

List of Illustrations

Section 2. COMPUTING MACHINERY (HARDWARE)

Example	
2-1	Digitizer
2-2	Coding Form
2-3	IBM Card
2-4	Line Printer
2-5	Tabular Output from Printer
2-6	Graphic Output from Printer
2-7	Pen Plotter
2-8	Cathode Ray Tube

Section 3. COMPUTER PROGRAMS (SOFTWARE) AND TYPES OF MAPS

Example	
3-1	Contour Map (SYMAP)
3-2	Conformant Map (SYMAP)
3-3	Interval Dot Map (SYMAP)
3-4	Symbol Map-Small Circles (SYMAP)
3-5	Symbol Map-Large Circles (SYMAP)
3-6	Contour Surface (SYMVU)
3-7	Conformant Surface (SYMVU)
3-8	Four-Dimensional Map (SYMAP)
3-9	Three-Dimensional Histogram
3-10	Single Character Grid Cells (GRID)
3-11	Multiple Character Grid Cells (GRID)
3-12	Conformant Map (Calform)

1. INTRODUCTION

The term "computer graphics"--as used in this report--refers primarily to the nature of the data being manipulated and displayed rather than to the computer equipment which may be used. As a part of this definition of the term computer graphics, it is assumed that the data being displayed has a spatial dimension.

It may be:

(i) one dimensional-as on a bar chart;

(ii) two dimensional-as on a map or scatter diagram; or,

(iii) three dimensional-as on an architectural object or statistical surface.

Such data--especially when a geographic distribution[1] is involved--is particularly relevant to studies concerning air pollution since the location or spatial pattern of a given variable may be as significant as its magnitude or quality.

1.1. CRITERIA FOR COMPUTER MAPPING

Although maps as such are not a new phenomenon, their preparation with the aid of a computer is a fairly recent development. At the present time any conceivable map could be prepared on a computer; however, not every map should be. In evaluating the possible use of computer mapping techniques for a given application, a user should consider the characteristics of the maps desired and data to be mapped.

1.1.1. Map Characteristics

Computer mapping can be a significant asset as an analytical tool during a study. This is in addition to the traditional role as a descriptive tool for displaying the end results of a study. Often maps can be useful for testing and evaluating alternative hypotheses and assumptions during the course of a study.

The mapping applications most likely to benefit from the use of computers have one or more of the following characteristics:

-- a large number of maps is needed at one time, or maps for a given area are needed on a recurring basis over a period of time;

-- flexibility as to size, content, scale and data manipulation are important;

-- uniformity of map appearance for many different types of data is desired;

-- the time required to prepare a number of maps is to be minimized; and,

-- highly accurate (though not necessarily precise) maps are needed.

1.1.2. Data Characteristics

Obviously, inaccurate or incomplete data should not be used, whether maps are prepared by hand or by computer.

This is particularly true of computer maps because many people place unwarranted faith in anything produced by a computer.

Computer mapping of the data is appropriate if some or all of the following characteristics are met:

-- large amounts of data are to be mapped;

-- the data is in machine readable form or can be readily converted;

-- the data contains geographic codes for the zones to be mapped or includes basic record identifiers which can be used to aggregate the data into zones;

-- possible sources of human error in handling the data and preparing the maps are to be minimized;

-- the data to be mapped measures a dynamic rather than a static phenomenon and will be updated to provide measures of change over time, or time series data is already available; and,

-- accurate data is available.

1.2. PRINCIPLES OF OPERATION

The preparation of a computer map requires three steps: (i) input--providing the computer with data in a form acceptable to the machine; (ii) processing--the manipulation of the data and preparation of the map within the computer's memory; and, (iii) output--the actual display of a map by the computer.

1.2.1. Input

Data to be displayed as a map must be available in a form which is compatible with the computer equipment (hardware) and the programs (software) to be used. This requires that the computer be able to "identify" every possible location on the surface of the map it is to prepare. The most common method of identification is a grid coordinate--or x-y reference--system which the computer can refer to whenever it needs to determine the location of a point--or the location of one point in relation to all other points on the map. Therefore, all data to be mapped should be assigned x and y coordinates or some similar system of locational identifiers.[2]

1.2.2. Processing

Once supplied with data for a given study area, a computer is capable of performing numerous arithmetic and logical operations upon this data. Calculations may be carried out on the locational as well as the quantitative aspects of the data. This might include calculations based upon distance, direction and value.[3] Whether the value to be mapped is part of the input or is derived by the computer, the resulting map is constructed on the basis of grid-coordinate locations and numerical values associated with those coordinate locations.

1.2.3. Output

Computer maps can graphically display a wide variety of data in many forms. The maps may display lines, tones, letters, numbers, symbols, or colors, or they may be

three-dimensional, with heights shown above or below a reference plane. The same results can certainly be achieved manually, but a computer calculates more rapidly, economically, and accurately than a man can.[4]

1.3 CURRENT AND FUTURE APPLICATIONS

Existing applications of computer mapping fall readily into two groups: (i) military and aerospace applications, and (ii) other governmental and commercial applications. The more sophisticated applications of the military and aerospace group are related to their greater financial resources.

1.3.1. Military and Aerospace Applications

Relatively little is known about the most advanced of today's computer graphic systems developed by and for the military since they are classified. It is likely, however, that these systems include the capability of displaying on a television screen such information as location, class, destination and cargo of any naval vessel.[5]

The great majority of military systems use the cathode ray tube (CRT)--a television-like display--as an output device so that the operator may "communicate" with the computer and in effect continuously interrogate the data file. The operator may immediately make new inquiries based on prior responses by the computer, rather than submitting a problem or request to the computer and waiting for several hours or days for the reply.

At a number of military installations[6] plotters are used to recreate all or a part of a map which has previously been read--using a digitizer--and stored on magnetic tape. At the time that the map is redrawn by the computer, it is possible to alter the map size or projection and also to add to the final map--from a separate source--additional data which relates to the area being drawn.

A particularly useful device that is used by the Army and Air Force is capable of automatically reading or digitizing the features of a map and storing the data on magnetic tape. Such a device eliminates the task of manually digitizing the data and greatly reduces both the time required and the possibility of human error.

1.3.2. Costs Involved

In addition to the extremely costly equipment required, systems such as the military ones described obviously need extensive data collection, storage and retrieval operations if they are to allow the user to communicate with the computer. Even if funds for equipment were available to governmental agencies other than the defense and aerospace group, additional funds would be needed to establish an adequate data base for use with a computer mapping system. The initial cost of constructing such a data base would in most cases be greater than the cost of the equipment required to manipulate and store the data.

The programs required to operate the system and to process the data often cost as much or more than the equipment. The critical shortage of qualified programmers

is a serious problem at most computer installations. However, this problem may be lessened by acquiring and adapting existing programs previously prepared by other organizations, rather than writing all new programs.

There is also a very real user cost involved in developing an explicit statement of what tasks are to be performed on the computer and in relating the individual tasks to each other and to the total system.

1.3.3. Other Current Applications

Non-military applications have included the automatic recording and reproduction of maps for atlases and for natural resource data.[7] Metropolitan transportation studies have often included maps of transportation networks showing such features as street capacity, actual and projected traffic volumes, origins and distinations of trips, and trips categorized by purpose or type of vehicle. Nearly all of these maps have been drawn by pen plotters.

Line printers have also been used to prepare computer maps, chiefly by organizations with a definite need for a computer mapping capability within a limited budget. The users have included urban planners, geographers, police administrators, transportation engineers, recreation analysts, and archeologists. One of the earliest applications was in the preparation of a Community Renewal Program; in this, as in many other applications, the data mapped described some of the socio-economic characteristics of an urban area.

Mapping projects have dealt successfully with large areas such as metropolitan regions, states and even countries, and with small areas--in one instance, with the structure of a living cell. The variables mapped have been as diverse as the study areas:

-- population and housing characteristics from the U.S. census,

-- subdivision activity from plat records,

-- natural resource data from aerial photographs,

-- land use from field surveys,

-- property values from assessment records,

-- cultural artifacts from archeological explorations,

-- law enforcement data from police records,

-- recreation demand from state park records.

Clearly, given a very large budget, any government agency could develop applications requiring the most sophisticated computer systems. The above mapping projects, however, involved minimal expense and minimal computer equipment.

1.3.4. Future Applications

Within the next few years, the use of computer mapping is likely to increase enormously, both in quantity and in the diversity of applications. Because of the rapid growth in the use of computers, costs for a given level of capability have been dropping each year and will continue

to do so. At the same time, the capability, speed, and especially the memory capacity of computers have been increasing.

Within the next fifteen years, the equipment and knowledge required to operate and benefit from what are presently the most sophisticated systems will perhaps be within the reach of most governmental agencies. Important changes which would help to make this possible are the sharing of the same basic equipment and data by many users.[8]

1.4. AIR POLLUTION CONSIDERATIONS

Most air pollution data is gathered for specific geographic locations. Air quality is measured at monitoring stations or modeled for points on an x-y reference grid. Emissions data may be determined at a point--such as for a power plant--on a line--such as for motor vehicles--or over an area--such as for residential space heating. Of importance in this data gathering is the use of a common coordinate system--Latitude/Longitude, Universal Transverse Mercator (UTM), or State Plane Coordinate System (SPC)--for all air pollution data.

The usefulness of maps--and computer mapping--in air pollution studies has only recently become apparent as data have been gathered with an adequate geographic distribution over a region. Until the last few years air quality was monitored at only one or two locations in most metropolitan areas. Similarily, when emissions were estimated for an area, they were rarely allocated spatially.

Air pollution data fulfills most of the characteristics for computer mapping outlined in the first section. There is a need for accuracy, a large number of maps over time, uniformity of appearance, and--in the near future--rapid production of maps, particularly for telemetered data and emergency control systems. Computer mapping of most air pollution data is appropriate because of the large amounts of data--often in machine readable form--to be mapped and the continuing need for data display and comparison over time.

There are no special considerations for the input, processing, and output in regard to air pollution maps; questions of data input procedures and the types of maps useful are discussed in sections 3 and 4. In the future Cathode Ray Tubes will probably be very useful for the display of telemetered data and for emergency control systems, but the line printer will continue to be the most useful output device for the majority of applications.

The high costs of equipment, programs, and data can be reduced by the use of shared facilities and common data transmission and storage methods. To do this an expanded national air sampling network is needed with standardized data collecting, telemetering, handling, and display. The national inventory of air quality and emissions data in Cincinnati is accomplishing much of this standardization and centralization.

2. COMPUTING MACHINERY (HARDWARE)

The computing machinery, or hardware, involved in producing computer maps and allied displays is composed of:

(i) the computer itself, including the central processor and the peripheral input/output equipment;

(ii) The input devices, including keypunches and digitizers; and,

(iii) the output devices, including printers and plotters.

2.1. COMPUTERS

Computers differ greatly among manufacturers and models in capability, size, cost and availability. Although several programs for the production of computer maps have been written for the smaller computers (e.g., IBM 1401), more sophisticated and versatile map displays require larger, more powerful--and more expensive--computers (e.g., IBM 7094, IBM 360/50, CDC 3600). The most elaborate computer graphic systems--such as those used by the defense and aerospace agencies--may be used solely for the preparation of graphic output. However, a computer will usually be used only a small part of the time to generate graphic displays, and the bulk of the time for other, more common applications.

It is always difficult to specify the computer capability required for data processing, since programs have been written for small, medium, and large computers. However, it is safe to say that size will become more

critical as the state of software programs advances. Increased file manipulation and statistical analyses prior to mapping, combined with more versatile display systems, will necessitate the use of larger, more powerful equipment.

Time sharing systems, remote batch entry, and interactive graphic systems--light pen attachments for cathode ray tube scopes--all offer possibilities. Such innovations, however, will only be available on larger computers and at a few locations in the foreseeable future.

Computer mapping programs such as SYMAP and SYMVU, in their present form, require larger computers--in terms of both the size of the central processor and the complexity of the peripheral input and output equipment.[9]

2.2. INPUT DEVICES

Information must be converted from a visual or a numerical form to a coding sheet. The numbers that have been recorded on the coding sheet are then keypunched onto cards or paper tape. Often these two steps are performed by different people increasing the chance for error.

The use of digitizing equipment[10] to define the locations of points or boundaries eliminates the necessity of recording the numeric codes on paper and the subsequent keypunching. Although digitizing is a more efficient procedure, it requires a higher level of sophistication, especially with regard to the following aspects:

(i) the boundaries must be recorded on highly accurate drawings. The preparation of these drawings may be time-consuming and expensive;

(ii) the digitizing hardware and recording equipment requires a reasonably high capital investment, which can only show a return with continued use; and,

(iii) experience has shown that unless highly skilled and very patient personnel are used, this procedure can generate errors which are time consuming to correct.

Digitizing procedures will be efficient in a study where large amounts of data are to be collected and there will be a continuing need which permits the purchase of the necessary equipment and training of personnel. For short term studies, the procedures will probably not be profitable.

Some advanced and sophisticated forms of remote sensing can record data directly in a machine-readable form. Others prepare an original record such as a photoplate which can be scanned by a sensor to record the desired information. These forms of directly recorded information often contain only an implied locational key. When this is the case, the data must be converted so as to contain a usable set of locational identifiers, compatible with other information.

Digitizing equipment of the type referred to above usually involves a drafting table, a mechanical arm with an attached pointer that can be moved to any point on the table, a keypunch, plus the electronic controls required to link the various devices together. Such a device is

illustrated in Example 2-1. A map or drawing is placed upon the drafting table and the pointer device located at the point whose x-y coordinates are to be recorded. The operator then pushes a button and a keypunched card is generated automatically. The operator then moves the pointer to the next point and repeats this operation. Up to 1000 points per hour may be digitized using such a device.

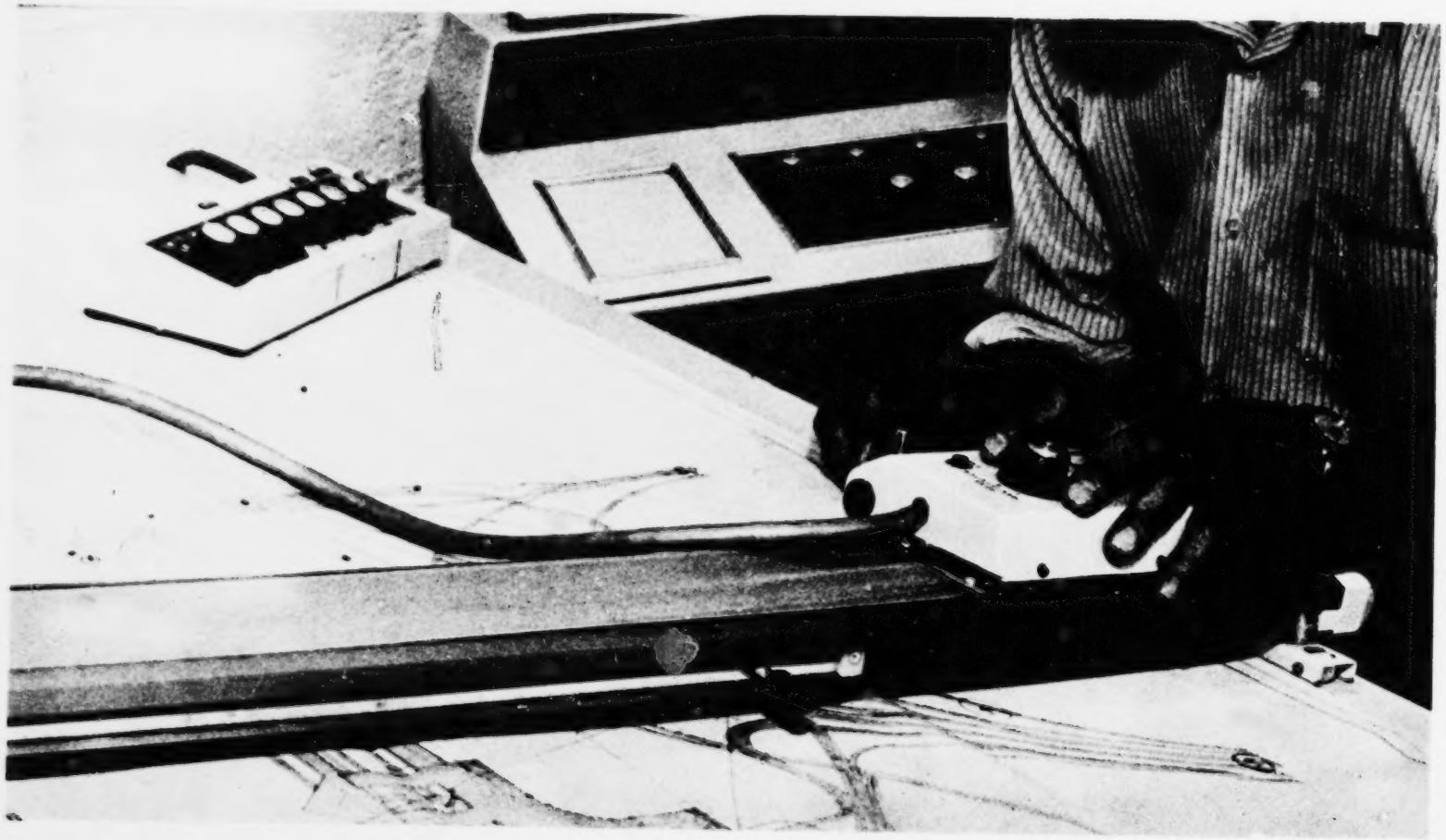
Digitizing devices are sometimes referred to as "inverse plotters" because rather than drawing or "plotting" information they are used to capture or record information which has previously been drawn or plotted by a human draftsman.

Example 2-2 shows a coding form of the type which might be used to record data. The contents of each row on the form would be keypunched onto one IBM card. Example 2-3 shows such a card.

Each character or digit occupies one column of the card. For a single digit (if a numeral) the corresponding number printed on the card is punched out in the appropriate column by use of a key punch machine. Multi-digit numerals occupy the appropriate number of adjacent columns. Letters of the alphabet plus various punctuation marks and special characters are recorded by punching out more than one numeral in a given column.

2.3. OUTPUT DEVICES

Graphics can be produced in some form on any of the many available computer output devices. However, the accessibility and cost of the equipment varies a great



Example 2-1
Digitizer

PROGRAM

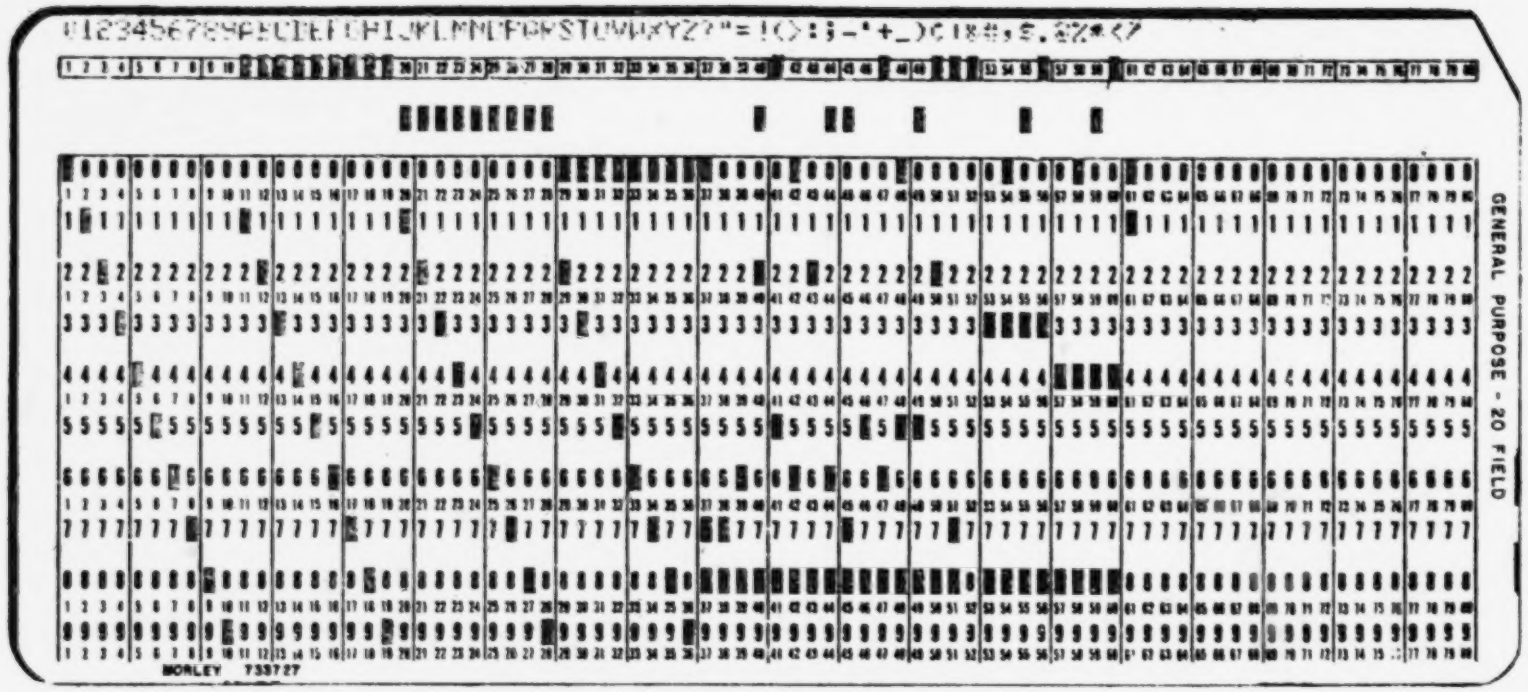
PROGRAMMER | DATE | PUNCHING INSTRUCTIONS | GRAPHIC / PUNCH

COMM. | STATEMENT NUMBER | CONT. | FORTRAN STATEMENT | IDENTIFICATION SEQUENCE

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 5 ... 77 78 79 80

A B C D E F G H I J K L M N Ø P Q R S T U V W X Y Z 0 1 2 3 4 5 6 7 8 9 + - / = * ( ) . , $

Example 2-2
Coding Form



Example 2-3
IBM Card

deal, as does the appearance of the graphics. The basic types of output equipment are the line printers and the mechanical or cathode ray tube plotters.

2.3.1. Printers

There are two basic types of printers: character printers and line printers. Character printers resemble typewriters and print only one character at a stroke. Line printers print a line of up to 140 characters at a stroke and can print 10 strokes per second.

Line printers are an integral part of almost every computer installation, making it a convenient and economical output device for computer maps. Line printers were not originally designed to print maps or other graphic displays, although they are able to produce a wide variety of computer output. They cannot draw continuous lines, but may be used to roughly approximate lines by a series of print characters.(11)

In addition, a map prepared in this manner can represent the location of a point with a precision of only one half the size of a print character. As a result, line printers are less suited than plotters for displaying line drawings; their lines may be accurately shown, but they are not precise.

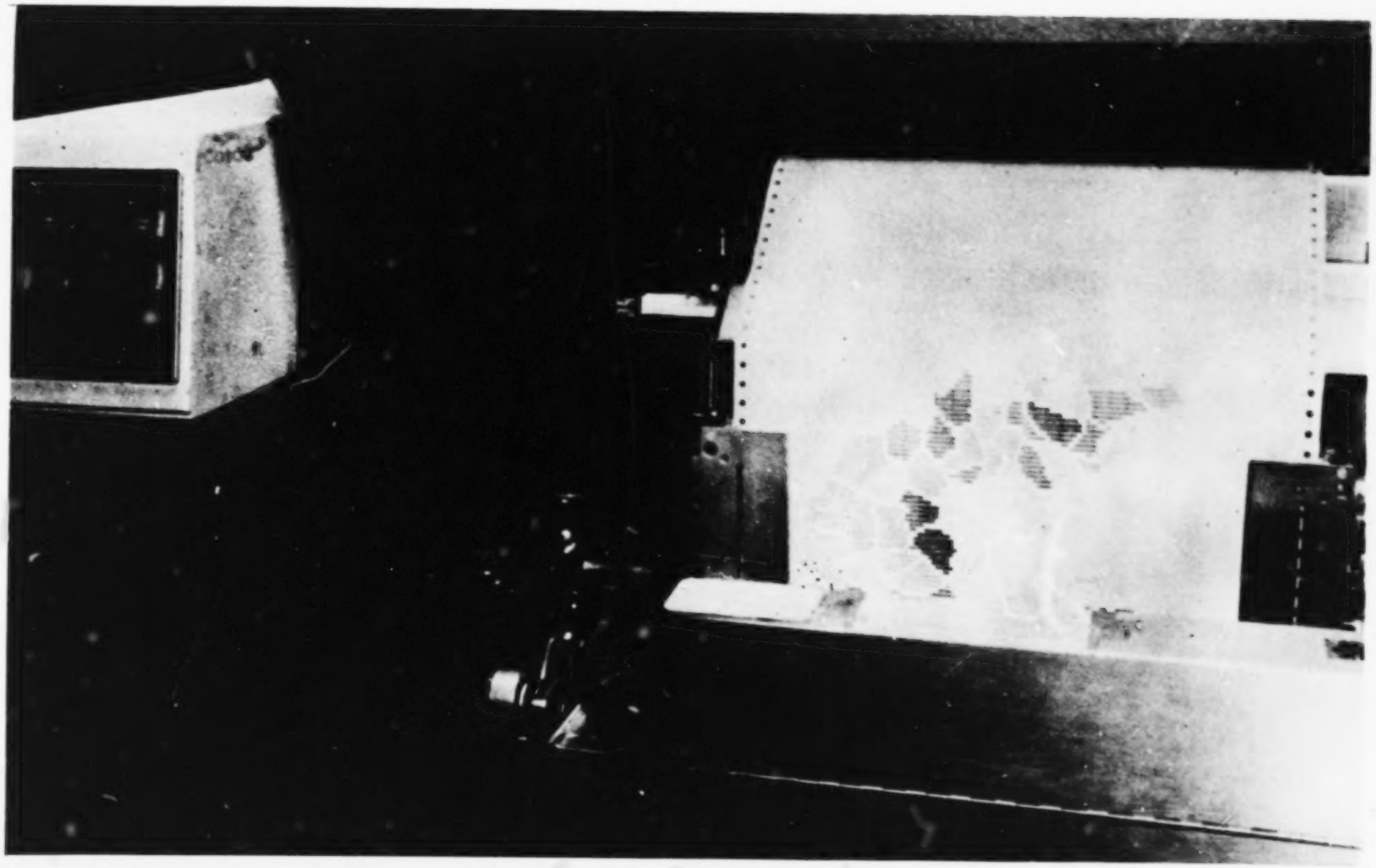
In the case of zone or area symbolism--such as that used on a contour map--a line printer is capable of providing a suitable end product--superior to that produced by line plotters in some ways. Tonal variations or

shading are easily prepared on a map by printing characters--or multiples of characters--at various print locations. Although the only characters available for use are those which appear on a given print chain, overprinting provides a wealth of possible symbols and tones.

Since line printers are designed to print alphabetic and numeric characters, they are obviously capable of printing words and numbers anywhere on the map surface. The letters are always printed horizontally--as they would be on a typed page--and cannot be rotated about their midpoint, as they can be when drawn by pen or cathode ray tube plotters. The drawbacks include the strict limitation of the graphic output to print characters--single or multiple--and to specified placement of characters at locations.[12] It is possible to obtain special print chains which have symbols more appropriate for graphic display. Such symbolism would eliminate the need for overprinting--with a consequent increase in speed and economy. Line printers to be used for mapping should also be adjusted to permit a line spacing of 1/10 inch which would make square print characters possible. Adjustments of the types recommended can be made at a cost which would be trivial[13] --when prorated over the many maps that could be produced.

Printer programs are also available which will generate plates for the offset printing of colored map displays.

A common line printer is shown in Example 2-4. Such a device produces an entire line of up to 140 characters each time that it prints. Up to 1,000 such lines



Example 2-4
Line Printer

can be generated per minute. As a result printed output prepared by such a device is produced at an extremely rapid rate and at a very nominal cost. An example of output produced by a line printer is shown in Example 2-5. In addition to the normal tabular output, printers may also be used to generate graphic displays such as that shown in Example 2-6.

2.3.2. Mechanical Plotters

There are two types of plotters: mechanical plotters, operating either on a drum or a flatbed[14], and electronic plotters, utilizing cathode ray tubes for image generation.[15] Plotter operations are either incremental (digital control, digital drive) or digital to analog (digital control, analog drive).

The incremental line plotters move their pens in small steps of 0.01 inches in the horizontal or vertical direction. The result is a fairly smooth line, although curves do take on a slightly sawtoothed appearance. The analog or drafting plotters, on the other hand, can continuously change direction and thus give smooth, straight or curved lines. Such lines are very precise, with an accuracy of up to 0.001 of an inch.

Plotters can draw letters, numbers and symbols as well as lines. Since drawing characters rather than lines is time-consuming, some plotters include a small print head capable of printing one character at a time. Plotters with several pens located on a turret can draw maps in color; a single pen plotter can also make colored maps, but this involves changing the pen for each new color.

A-OUTLINE

VERTEX	DOWN	ACROSS
	ISLAND	1
(1)	0.06	4.15
(2)	0.66	5.79
(3)	0.82	6.29
(4)	0.83	7.34
(5)	1.32	8.55
(6)	1.43	9.22
(7)	2.13	11.22
(8)	2.60	11.87
(9)	3.94	10.49
(10)	4.30	10.27
(11)	4.80	10.45
(12)	6.17	11.60
(13)	6.37	11.61
(14)	6.66	10.89
(15)	6.83	10.04
(16)	6.39	8.83
(17)	6.43	8.76
(18)	6.45	8.21
(19)	6.34	7.74
(20)	6.19	7.36
(21)	6.19	5.78
(22)	6.57	4.36
(23)	6.39	4.00
(24)	7.18	3.19
(25)	7.18	2.60
(26)	5.07	1.80
(27)	5.30	1.60
(28)	5.29	1.40
(29)	3.92	0.41
(30)	3.28	0.41
(31)	2.57	1.18
(32)	2.62	1.28
(33)	2.33	1.50
(34)	1.72	1.50
(35)	1.31	1.98
(36)	1.04	2.82
(37)	1.03	3.31

AREA= 54.09 CENTER=(3.82, 6.04)

Example 2-5
Tabular Output from Printer

DATA VALUE EXTREMES ARE 0.0 98.41

ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL
("MAXIMUM" INCLUDED IN HIGHEST LEVEL ONLY)

	1	2	3	4	5
MINIMUM	0.0	28.81	53.41	70.34	80.13
MAXIMUM	28.81	53.41	70.34	80.13	100.00

PERCENTAGE OF TOTAL ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

1	2	3	4	5
28.81	24.60	16.93	9.79	19.87

FREQUENCY DISTRIBUTION OF DATA POINT VALUES IN EACH LEVEL

LEVEL	1	2	3	4	5
SYMBOLS	.	+	C	6	5

FIG. 6.

```
    1     2     3     4     5
 1  I..1..I  I++2++I  ICC3CCI  I664665I  I665566I
 2  I..1..I  I++2++I  ICC3CCI  I664665I  I665566I
 3  I..1..I  I++2++I  ICC3CCI  I664665I  I665566I
```

Example 2-6
Graphic Output from Printer

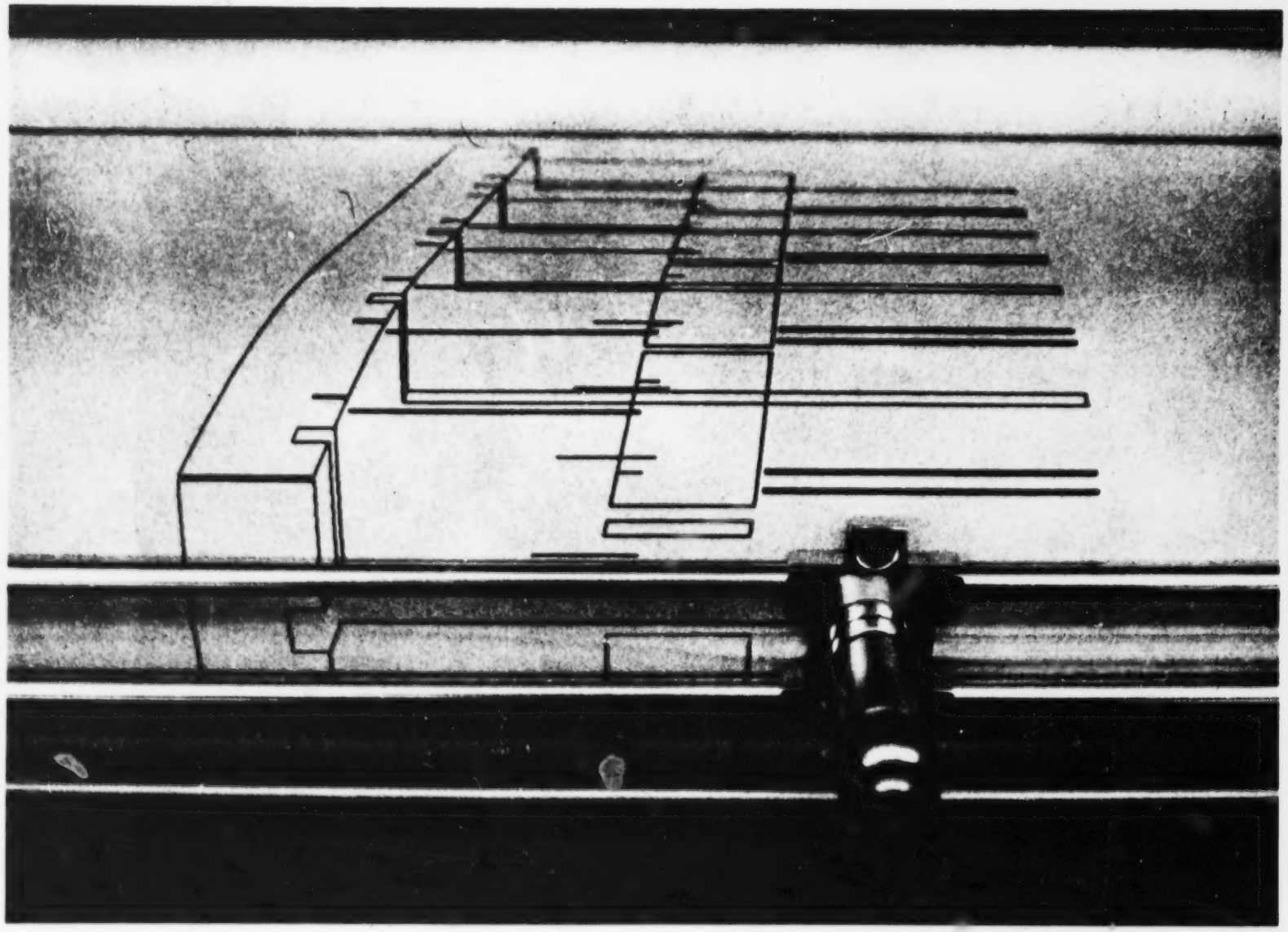
In general, the plotters are capable of providing very precise and attractive output. They are not available at every computer installation because of their cost and relatively infrequent use.

2.3.3. Cathode Ray Tubes

The last category of output equipment is the cathode ray tube (CRT). An image is displayed directly on the face of the tube with the screen size limiting the amount of information which can be displayed at any one time.[16] The screen is fairly well suited to displaying lines but not to showing areal patterns composed of symbols.[17]

At present, CRT equipment is quite expensive, relatively few units are available, and only a few programs for graphic display have been developed. However, all of these deficiencies are likely to be overcome in the future. The increased use of CRT equipment will be justified since they allow the user to communicate with the computer--to actually alter the image displayed and thus rapidly evaluate alternative displays. The technology already exists to couple the CRT with devices which can print on paper the image shown on the screen. The armed forces have been using the 'CRT System' for several years, but non-military experience in its use has been very limited.

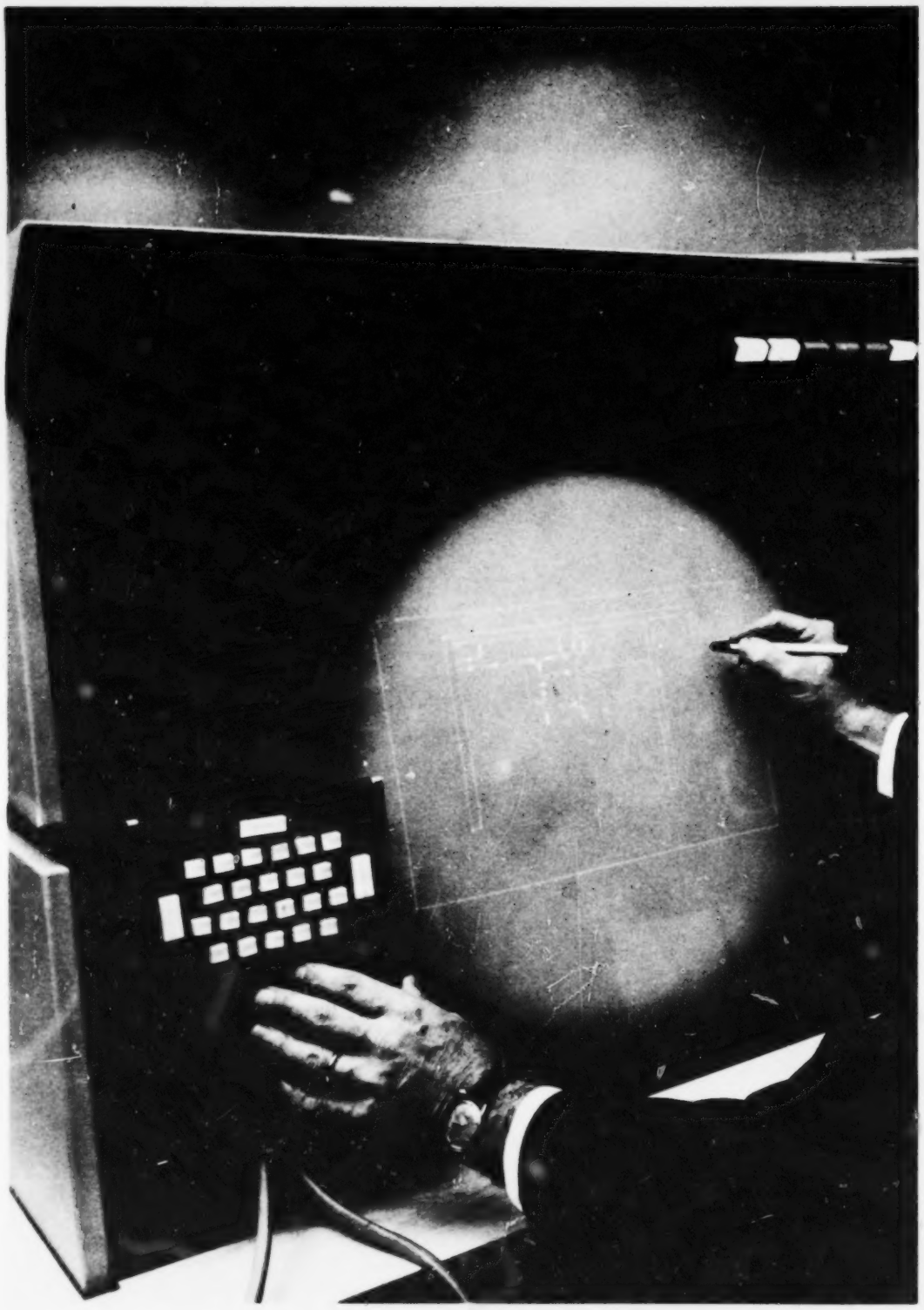
An example of a pen plotting device is shown in Example 2-7. The paper is on a continuous roll and capable of being moved forward or backward over the drum. A pen may be moved across the drum, from left to right in either a raised or lowered position. The move-



Example 2-7
Pen Plotter

ment of paper on the drum and the pen across the paper may occur essentially simutaneously.

Example 2-8 shows a Cathode Ray Tube. It is very similar in appearance to a television screen, and typically produces displays using a 1024 x 1024 matrix. Each cell of the matrix may be illuminated on the face of the tube at a given level of intensity. Some CRT devices allow for images to be drawn directly on the screen using a light pen or rand tablet. The user is able to achieve a high degree of man-machine interaction in the preparation and evaluation of graphic displays.



Example 2-8
Cathode Ray Tube

3. COMPUTER PROGRAMS (SOFTWARE) AND TYPES OF MAPS

It is possible to prepare many types of maps with the computer, but no single computer mapping program is capable of producing all of them. Therefore, the choice of the type of map desired for the graphic display often determines the program to be used. In this section, first, the types of maps and, second, the computer programs are discussed.

3.1. TYPES OF MAPS

The selection of the map type to be used should be based upon many considerations including assumptions about the geographic distribution of the data and its classification for purposes of display. Data which has been recorded for each of several zones will have a geographic distribution which can usually be thought of as either uniform or continuously variable. Values of a qualitative nature such as land use will typically be displayed in a manner which assigns a single category or land use type to each zone. Quantitative data--such as emission densities of air pollutants--are usually assumed to be continuously variable in their geographic distribution and should be so represented graphically.

The classification of data values for purposes of display represents a major decision. In order to graphically represent the spatial patterns in the data, it is necessary to assign a common symbol to those locations which have a similar value. For example, the data may be organized into several categories or classes, each of which has a specified range of values and an associated graphic symbol such as a color or tone of gray. Decisions of this type--i.e.,

the number of levels and the value range of each will have a profound effect upon the final graphic display.

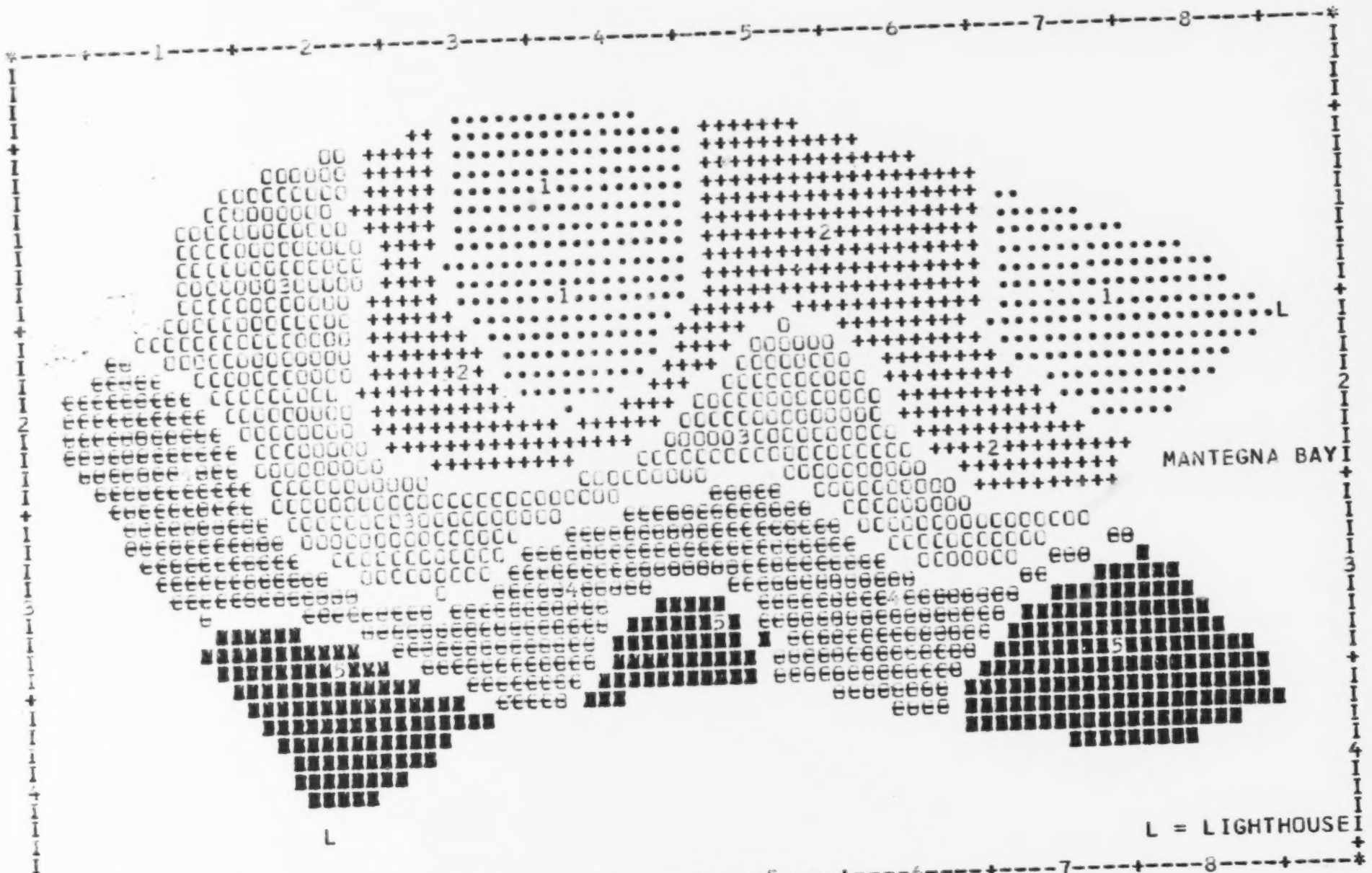
As an alternative to establishing such classes, the user may wish to display his information as incidence--rather than as zonal aggregates--where each occurence might be represented as a dot. If the data is not of a binary nature (present or absent) but rather specifies magnitude at each location, the information may be displayed as a third dimension or elevation above each data location.

A number of programs providing several alternative map types are discussed below.

3.1.1. Contour Maps

The continuous distribution of a phenomenon over a study area can be represented by a contour or isoline map. The phenomenon need not be topographic elevation, which is traditionally associated with 'contour' maps. Such maps may be thought of as two-dimensional, plan view representations of three-dimensional surfaces. Each contour line represents a line of constant elevation, or value, upon the surface. It is usually desirable to shade the areas between contour lines such that the intensity of the shading is a function of the elevation or "z" value. This shading helps to show the direction and intensity of the slope of the surface.

Example 3-1 shows a contour map produced by SYMAP for an hypothetical study area, Mantegna Bay. A surface was interpolated from the values at 14 data points. This surface was then sliced, horizontally, at four elevations and the areas between slices shaded.



Example 3-1
Contour Map (SYMAP)

3.1.2. Conformant Maps

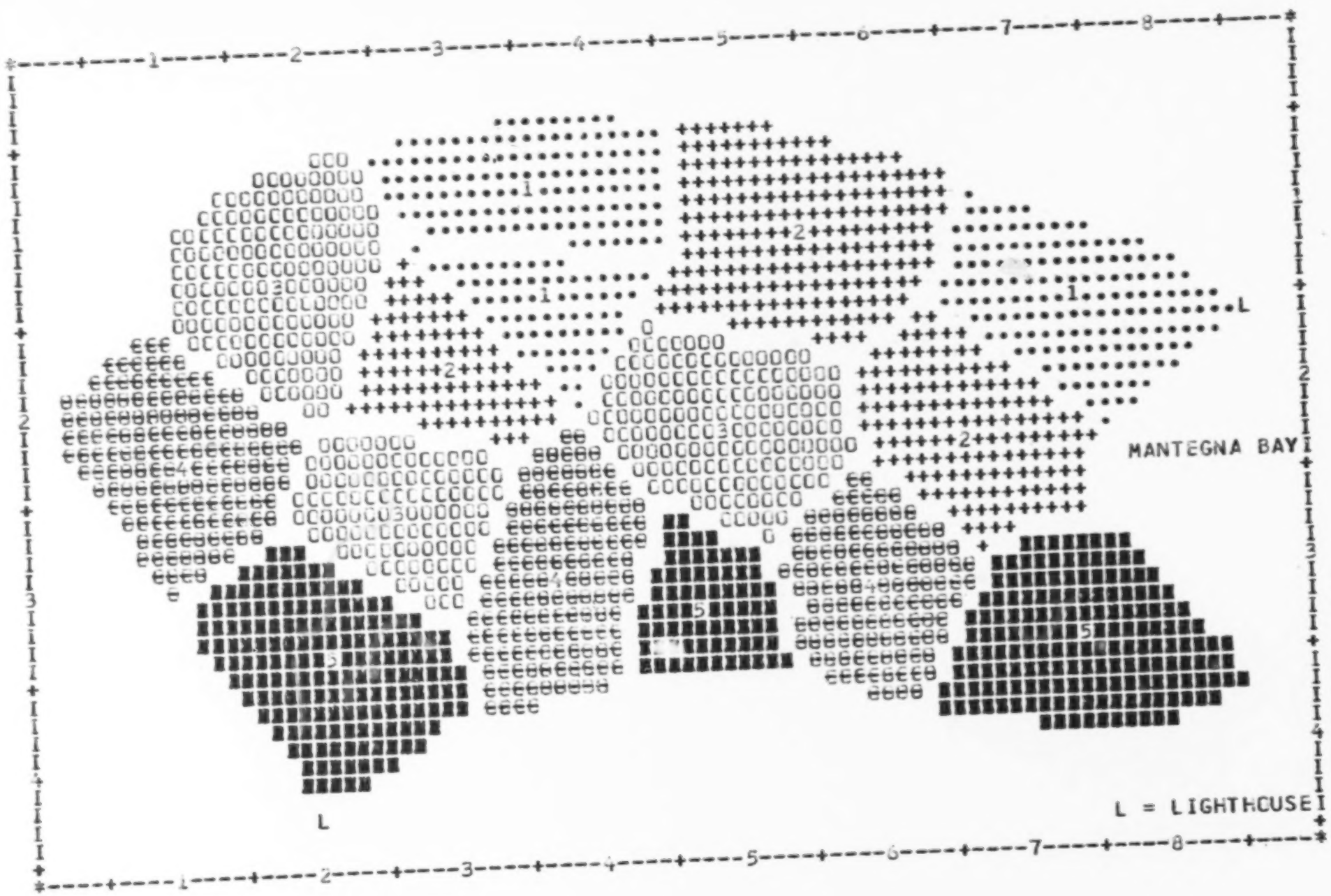
Conformant (or choropleth) maps represent data zones where a uniform distribution of the variable over each zone is assumed. Therefore, these maps are displayed with a uniform shading for each zone. The categories into which the value range may be divided are limited by the number of unique shades of gray the viewer is capable of differentiating among.[18] Generally, the user will wish to display far fewer than this limiting number.

Example 3-2 shows a conformant map for the same information displayed in Example 3-1. The study area is divided into 14 zones; the same 14 values as were used in Example 3-1 are assigned to these zones. The value range is divided into five equal intervals and each zone shaded according to the position of the value in the total value range.

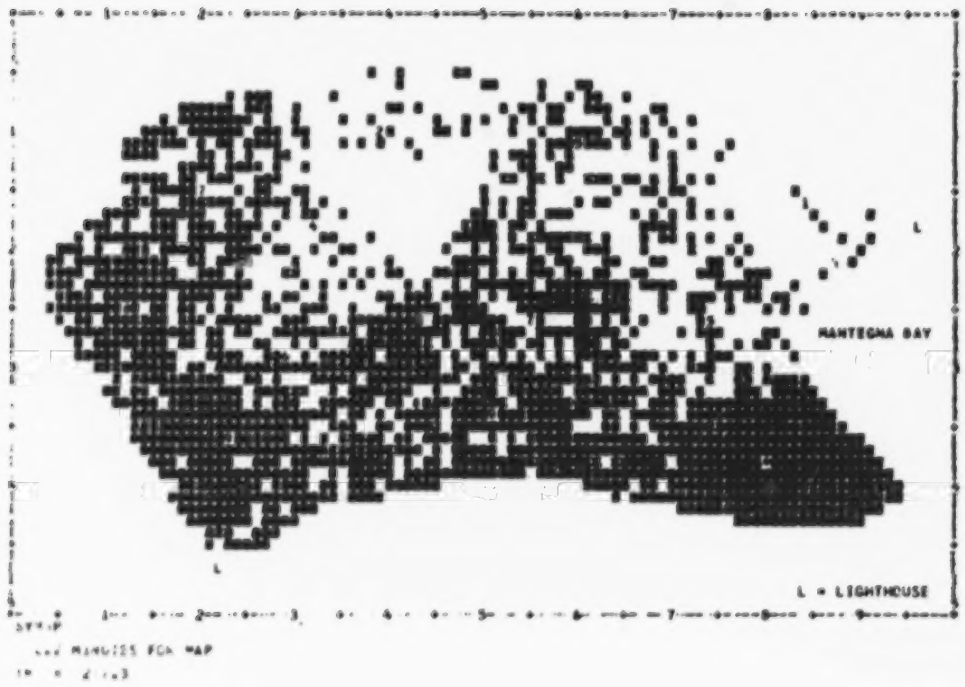
3.1.3. Dot Maps

Dot maps may be of two types: incidence or interval. For the first type, each dot represents the single occurrence of a phenomenon or incident at the precise location shown. An interval dot map, on the other hand, may represent more than one occurence as a single dot. The dots may be located in either a fixed or random pattern, but only within the given zone for which the data has been reported.

Example 3-3 shows an interval dot map for the same information shown in Example 3-2.



Example 3-2
Conformant Map (SYMAP)



Conformant Dot Map

Example 3-3
Interval Dot Map (SYMAP)

3.1.4. Symbol Maps

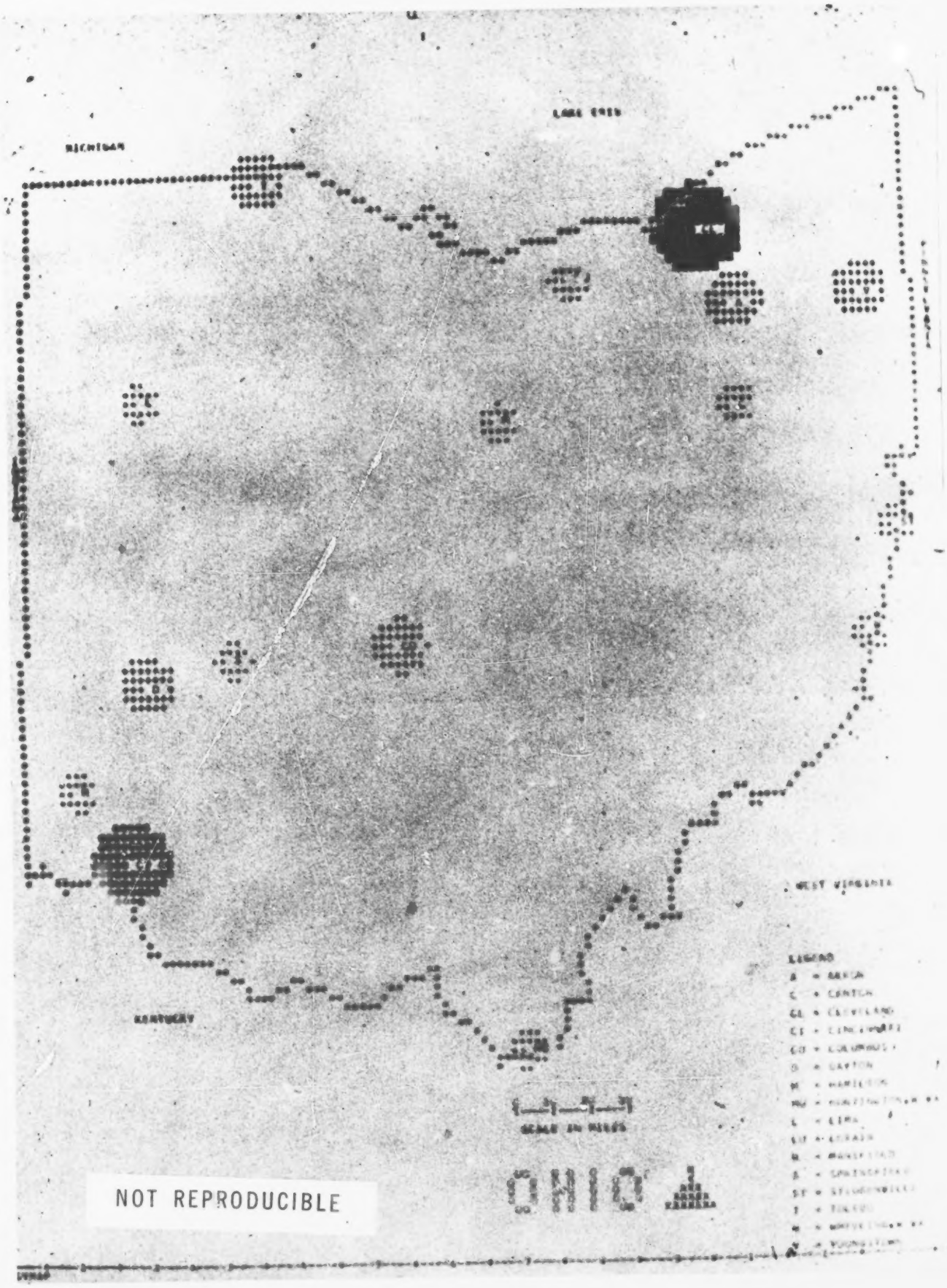
The value at a point or within a zone may be represented by a symbol which varies in size and/or tone as a function of the value. The most common form of such a map displays circles of varying size and tone--other forms include squares, and triangles. As an alternative to varying the size or tone of a single symbol, one may vary the number of times which the symbol occurs at each map location.

Examples 3-4 and 3-5 illustrate two symbol maps for population of Ohio cities. The data is identical in both cases but the circle size for a given value has been varied. This is equivalent to changing the value range for each gray scale on images of that type. The effect of increasing or decreasing the circle size is to emphasize the smaller or larger cities respectively.

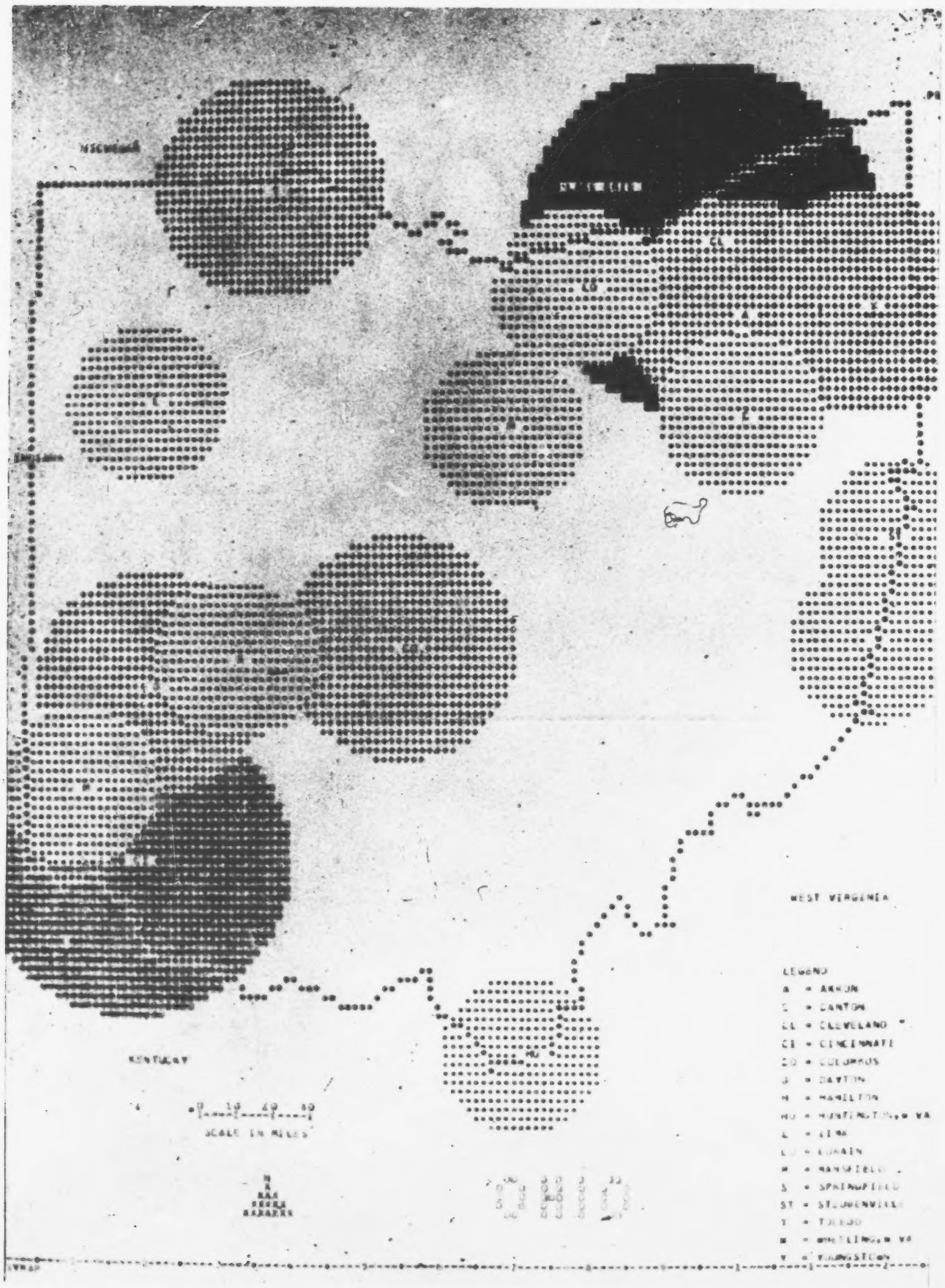
3.1.5. Three-Dimensional Maps

Three-dimensional maps represent the value at each location as a vertical height above a reference plane. For variables which have a continuous distribution over a study area, such a map would have the appearance of smooth, undulating surface. If the phenomenon is assumed to be uniformly distributed within data zones, the surface would have the appearance of a series of plateaus. The plateaus would retain the outline of the data zones, but the elevations would reflect the value within each zone.

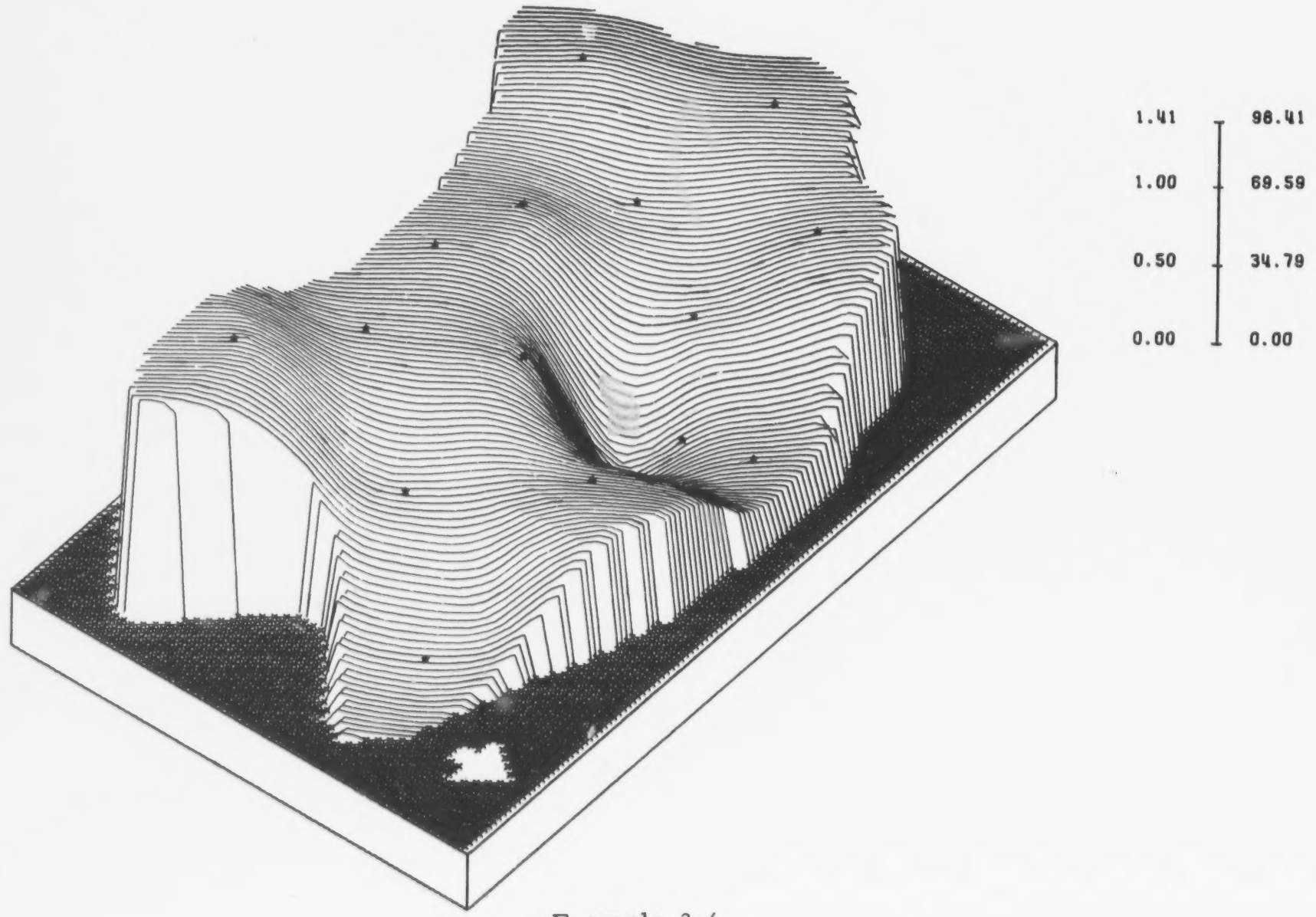
Example 3-6 shows the same surface displayed by the contour map of Example 3-1; Example 3-7 shows the con-



Example 3-4
Symbol Map-Small Circles (SYMAP)



Example 3-5
Symbol Map-Large Circles (SYMAP)

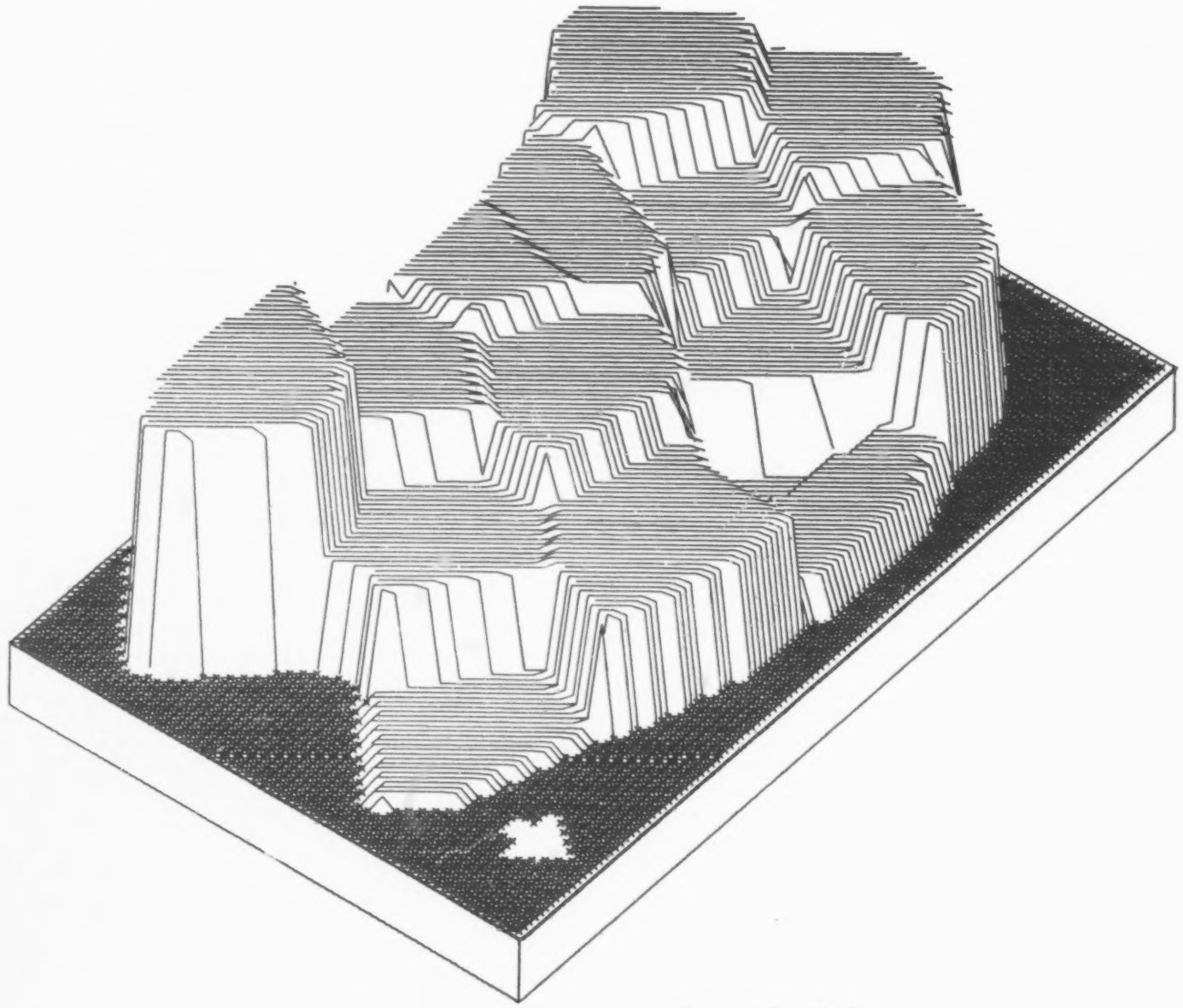


Example 3-6

FIGURE B: MANTEGNA BAY - CONTOUR

AZIMUTH = 231 ALTITUDE = 45
*WIDTH = 6.00 *HEIGHT = 2.00

* BEFORE FORESHORTENING 09/25/70



Example 3-7

FIGURE A: MANTEGNA BAY - CONFORMANT

AZIMUTH = 231 ALTITUDE = 45
*WIDTH = 6.00 *HEIGHT = 2.00

* BEFORE FORESHORTENING 09/25/70

formant zones, or plateaus, that were represented in Example 3-2. These two examples were produced by the SYMVU program.

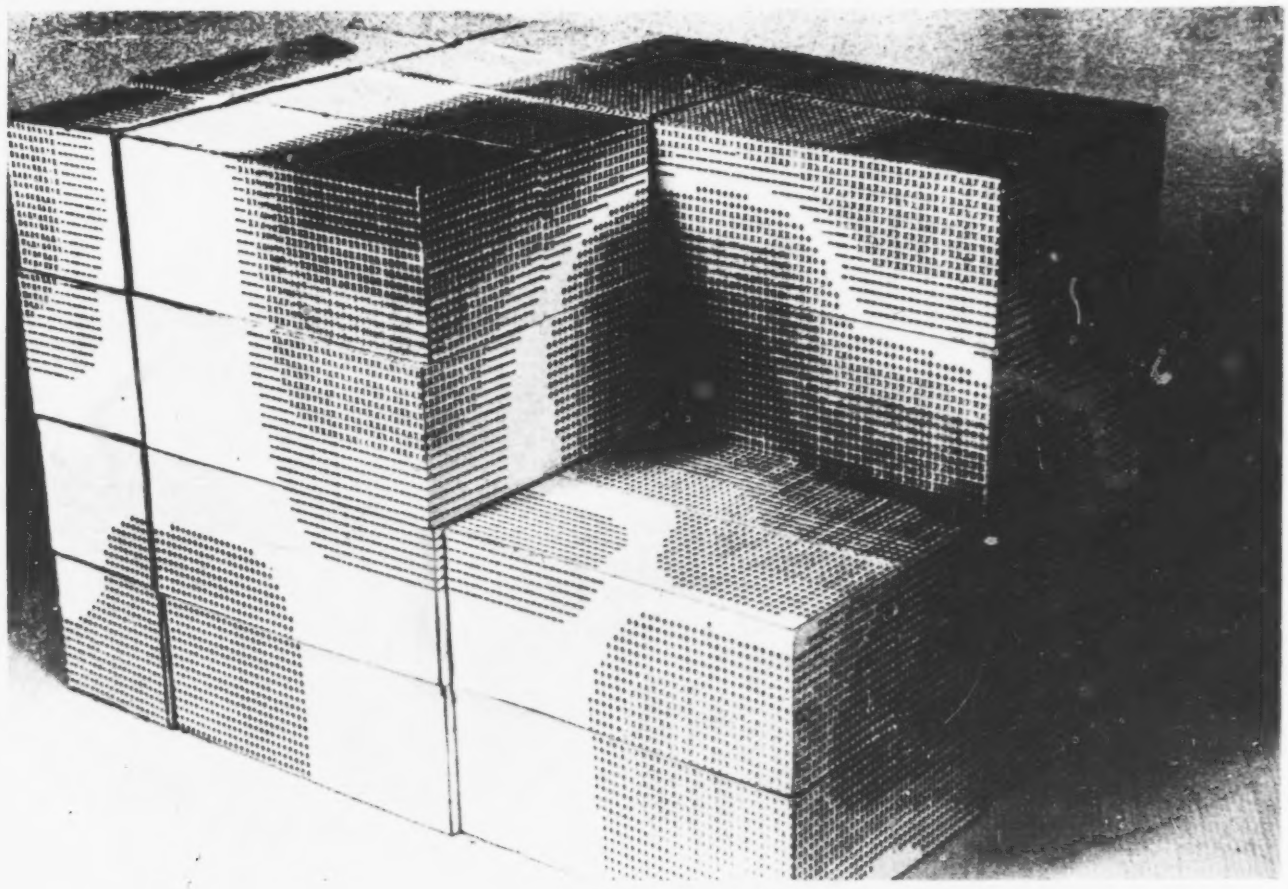
3.1.6. Four-Dimensional Maps

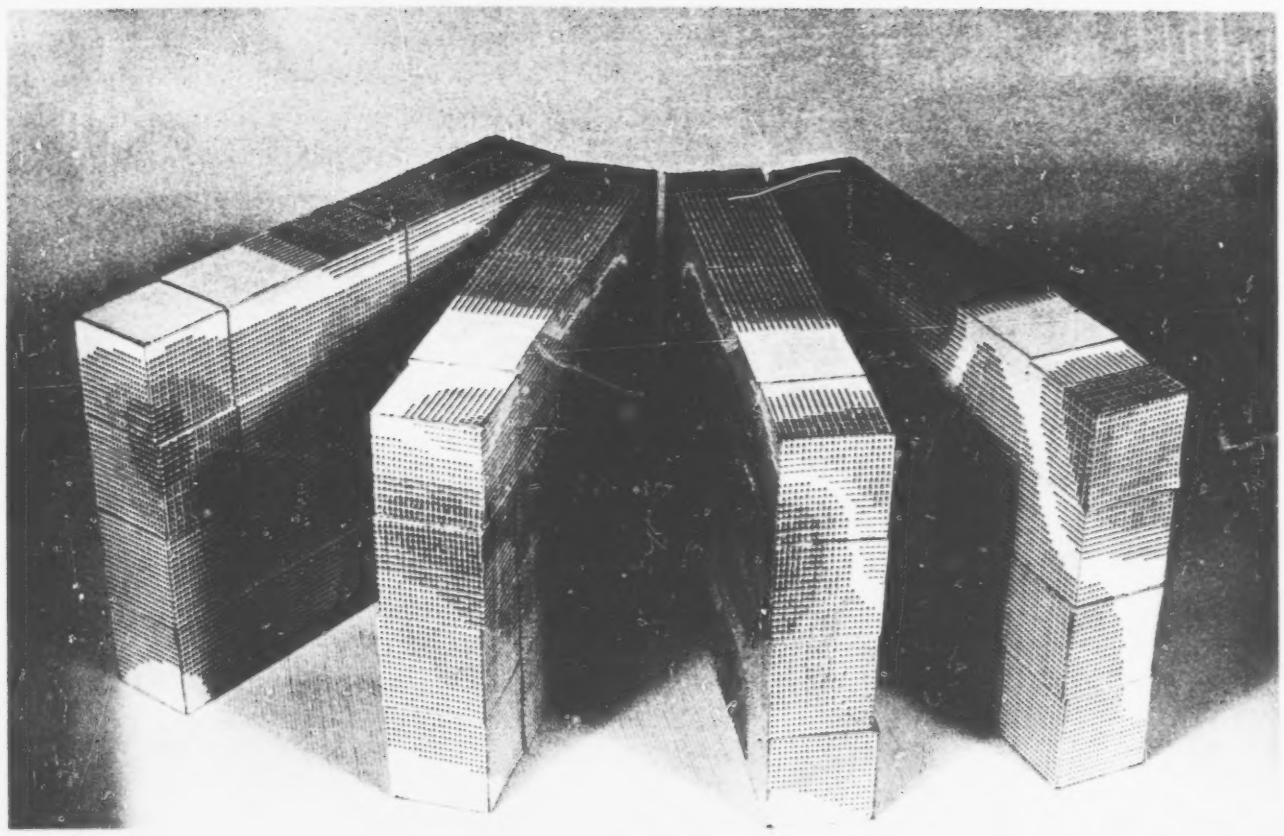
Four-dimensional maps represent the value of a variable in three-dimensional space. A model for this display technique was first constructed at the Laboratory in 1966. A series of eighteen twelve-inch SYMAPs, mounted in sections on a cubic assembly of twenty-seven smaller (4 inch) cubes, represents a three-dimensional study area. Data fluctuations are visible on the surface and can be seen internally by the removal of one or more of the smaller cubes.

Example 3-8 shows such a four-dimensional map for the velocity of ocean currents in a section of the Pacific Ocean.[19]

3.1.7. Color

The use of color has fairly specific but limited application in the preparation of thematic maps by computer. The added cost associated with such displays is generally not justified by their added value. The most practical method of preparing such displays requires several color separation plates. Each plate is made from a separate map;[20] the plates are then printed by a photo offset method. Specific uses for color displays include maps of multiple subjects, or qualitative data, and for showing gains and losses on the same maps.





Example 3-8
Four-Dimensional Map (SYMAP)

3.1.8. Non Geographic Mapping

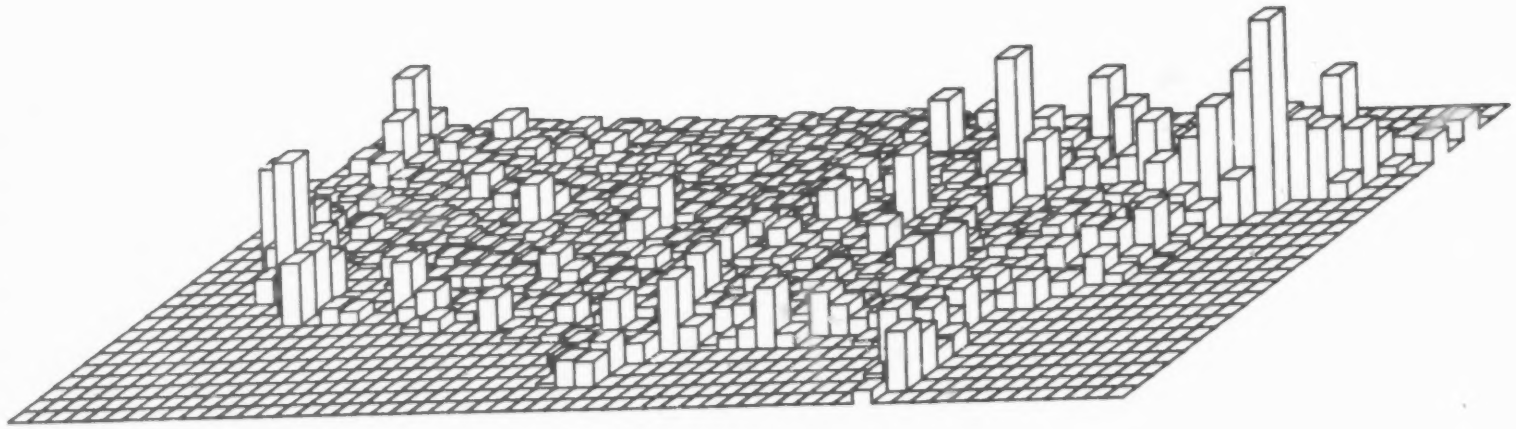
Although the above comments have dealt primarily with maps of geographic data, non-geographic graphic display is frequently quite valuable and is capable of being produced by the computer. Examples of such graphics include graphs, scatter diagrams of statistical data, bar charts, and pie diagrams.

Example 3-9 shows this type of display for geographic data, although the choice of the x and y axes is arbitrary. Population distribution is shown in 1960 for one degree quadrilaterals of latitude and longitude.[21]

3.2. COMPUTER PROGRAMS

Computer programs for producing maps and other graphic displays are usually written for a particular type of output device--such as a CRT, pen plotter or printer. The precision or texture of the output is a function of the output device and not of the program used. For example, the computer may accurately calculate x and y coordinates to eight significant digits. The line plotter would then produce lines connecting coordinates and having a tolerance of $\pm$ 0.005 inches On the other hand, the printer would display discrete symbols, representing the coordinates and the lines within 0.12 x 0.10 inch print positions.

The available computer mapping programs are, therefore, discussed in terms of the output devices used.



Example 3-9
Three-Dimensional Histogram

3.2.1. Plotter Programs

Pen plotters are relatively scarce; accordingly, plotter programs are uncommon and tend to be special-purpose in nature.[22] Programs for plotters are characteristically designed to locate precise values at precise locations. There are, for example, plotter programs which will reconstruct the features of a base map stored on tape, drawing street names, streets, building lots, rivers or topographic contours. Quantitative data from separate sources may be added, such as traffic volumes for each street segment, land value for each lot, or number of persons in each building. This quantitative data may appear as numerals drawn at the desired location on the map or as symbols representing data levels, or categories of data values.

Although the is olines produced by plotters and plotter programs are very precise and provide excellent final displays, it is often difficult to visualize the conceptual surface which they represent. Since all lines are identical in appearance, only the numerical values assigned to adjacent lines reveal whether the surface is rising or falling. It is possible for a pen plotter to fill the area between each of the contour lines with symbols, chosen to reflect the transition from high to low values or vice versa. Because of the great number of lines necessary for the shading, the time required to do so would substantially increase the cost of the map.

Cathode Ray Tube plotters operate with such speed that shading or other supplemental symbolism does not require a signficant amount of time, beyond that required for the basic line map. CRT plotters--unlike many pen plotters--cost less for the plotting time than for the computing time; however,

the preparation of the map by the computer is still as costly and time-consuming as ever, whatever the merits of the output device.

Programs prepared for plotting devices can generally be used on either a pen plotter or a CRT plotter with minor modifications. Plotter programs in general are capable of providing very precise and aesthetically pleasing output, but at greater cost than other means. The equipment and programs are usually the best choice for preparing a small number of maps for use in publications or public display.

3.2.2. Printer Programs

Printer programs locate standard print symbols and combinations thereof at pre-established print positions, rather than locating values precisely at a point on the map. For most applications where numerous maps are desired during the course of a study, the printer is the best choice because of its speed, its widespread use at almost every computer installation, and its low operating cost.

There are a variety of mapping programs which are designed to print, rather than plot, the output. One of the first was developed by Edgar Horwood at the University of Washington. This program has the capability of printing one or more numerical values at specified locations on the printout paper. A transparent overlay containing the boundaries of data zones is then superimposed on the printout. It is then possible to tell what numerical values fall within specific zones.

Another kind of mapping program was developed by the New York State Transportation Study. This program, Map 01,[23] can print one or more numerical values at specified locations. It can also be used to print a solid black character at each print position, similar in appearance to dot maps. Data value locations are assigned to a grid which has been superimposed on a map of the study area. A value assigned to any cell will fall within one of 20 data levels;[24] the number of "dots" appearing in the cell is equal to the level number for that cell. It is possible to print the numerical values in each grid cell, instead of the dots.

The R. L. Polk Company uses a program similar to Map 01 to produce computer maps. A base map is photographically enlarged or reduced until a chosen geographic area--usually a city block--corresponds in size to a single print position. The print characters used are single numerals, 0-9, which refer to pre-established value ranges for each data zone.

The most flexible printer program currently available was developed by Howard T. Fisher in 1963 at Northwestern University. Modifications and improvements over the past few years at the Laboratory for Computer Graphics and Spatial Analysis under the direction of Fisher and others have produced the present version of SYMAP. The SYMAP program requires the use of a relatively large computer such as an IBM 360/40. Although computers of this size may not always be available to local governmental agencies, they are common at large universities and business firms.

In summary, computer mapping programs written for line printers offer the user significant economical advantages. Although maps produced by plotter programs and equipment are usually superior, printer programs and equipment can be used to prepare numerous maps rapidly and relatively economically for analytical purposes during a study, as well as for display purposes at the end of the study.

3.2.3. SYMAP

SYMAP generates map displays showing the values of spatially distributed data, according to their actual geographic location on a base map. Combinations of standard computer printout symbols are used to print a black-to-white tone range which can correspond to a large-to-small data value range. Any mathematical computations can be made on the data, and options are provided for the rescaling of raw data and for variations in symbolism.

The SYMAP program[25] is written in Fortran IV. It can be used on a computer system equivalent to the IBM 360-40, or CDC 3600. Computer storage requirements are at least 32,000 words, or 128,000 bytes. Such systems are available in most commercial and university computation centers.

The program can produce conformant, contour, proximal, and dot maps. In conformant maps, predefined areas, such as census tracts, are filled with the symbol tone assigned to their data values (Example 3-2). Contour maps show the variation of the data over the whole study area, and the display is analogous to the contour bands used to show

changes of elevation on topographic maps (Example 3-1). Proximal maps use the nearest-neighbor principle, with any non-data position being given the value of its nearest data point. They are similar in appearance to conformant maps. The program will also produce dot maps, using a variety of symbols (Example 3-3).

To produce computer maps, a deck of punched cards must be prepared as input to the computer. This deck consists of a number of "packages," each covering a specific category of information. The most basic of these packages, with a brief explanation of their general purpose, are listed below. MAP ELECTIVES is always required; use of the other packages is determined by the type of map desired.

The MAP ELECTIVES package specifies, below the map, an appropriate title for the identification of each separate map that is to be run. It then instructs the computer to make each map by using certain electives, as required. The availability of these electives is a major feature of the SYMAP program and adds greatly to its flexibility. Some of the electives are:

(i) Size: used to specify, by inches and decimal fractions thereof, the overall dimensions of the rectangular map border surrounding the study area.

(ii) Content: enables the computer to change the scale of the map to be produced or to map a portion of the study area at the scale originally specified or any other scale desired.

(iii) Number of Levels: used to specify the number of levels and/or class intervals (not more than 12) into which the total range of the data is to be subdivided for mapping purposes.

(iv) Histogram Bars: used to suppress the normally-provided bars whose length shows the relative frequency distribution of the data.

(v) Text: used below the map to provide supplementary information.

(vi) Other Electives: deal with such subjects as Symbols, Missing Data, Maximum and Minimum Values, Borders, Scale, etc.

The OUTLINE package describes the outline of the study area, if nonrectangular, by specifying the coordinate locations of the outline vertices.

The DATA POINTS package is used to give the positions of the data points to which data are to be related, by specifying their coordinate locations. Data points may be either points for which data are available or centers of areas ("data zones") for which data are available. When warranted by the nature of the study, other "centers" may be used.

The OTOLEGENDS package causes supplementary information to appear on the face of the map by specifying their coordinate locations and content. Legends include borders, location names and numbers, the title block, etc.

The VALUES package assigns numerical data to the data points, by specifying the values involved. Raw data given to the computer may be related, manipulated, weighted, and aggregated in any manner desired.

Immediately ahead of the packages of cards described above, certain introductory cards (as required by the particular computing center being used) will need to be provided--together with a copy of the SYMAP program on cards or on tape.

3.2.4. SYMVU

The SYMVU program[26] provides the user with oblique views of a given surface by generating a magnetic tape for an incremental pen plotter. Although SYMVU was originally written in Fortran IV for the IBM 7094, it has since been converted for use on an IBM 360/40.

The user of SYMVU needs to supply a matrix of surface values, either on cards or magnetic tape (binary), and a set of parameters specifying the type of oblique view desired. The basic options available to the user are:

(i) the oblique view may be orthographic or in perspective;

(ii) the viewing angle may be varied in both the vertical (z) direction and horizontal (x, y) plane;

(iii) the lines that SYMVU generates to define the surface of the three-dimensional view may follow down the columns, across the rows, or through the diagonals of the input matrix;

(iv) the size is specified by width and height;

(v) treatment of the non-study area may consist of deleting it, darkening it, or setting it to a value of zero;

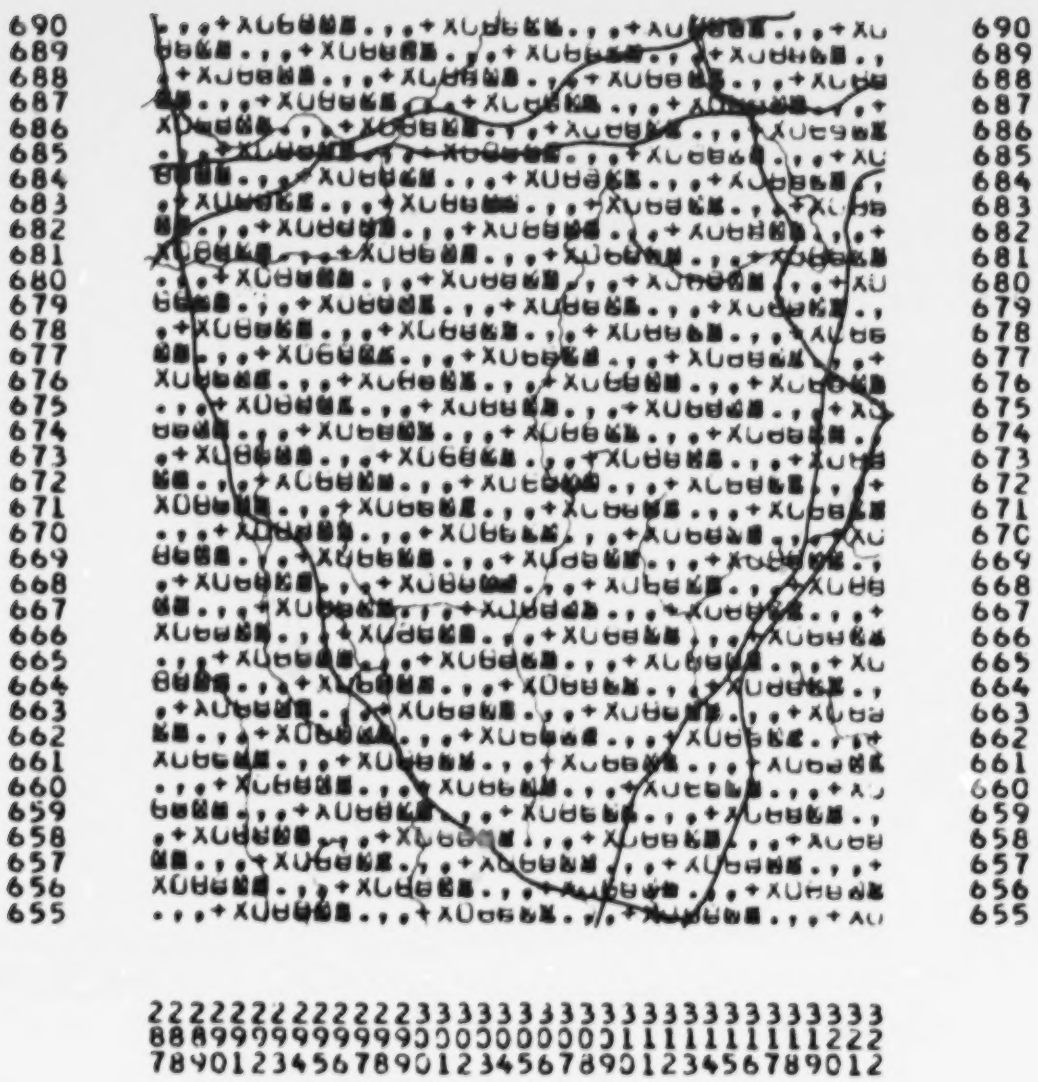
(vi) point symbols may be superimposed on the surface of the oblique views to denote data points, contour lines, rivers, a north arrow, etc.; and,

(vii) legends may also be added to locate various physical landmarks, indicate the scale, etc.

SYMVU is especially well suited to be run in conjunction with SYMAP or GRID, which will generate the necessary type of matrix values. The user then specifies only those options most appropriate to his needs. Sample SYMVU outputs are shown in Examples 3-6 and 3-7; these are identical to the surfaces displayed by SYMAP on Examples 3-1 and 3-2, respectively.

3.2.5. Other Mapping Programs

Two of the additional programs of particular interest for preparing computer maps are the "GRID" and "Calform" programs. The GRID program[27] is a modified version of SYMAP specifically designed for displaying data collected on the basis of geographic grid cells. Each grid cell may be shown on the computer-produced map as a single character or group of characters. An example of a single character per cell is shown in Example 3-10 and groups of characters per cell in Example 3-11. The program displays its data using a line printer but operates more efficiently and economically because of its special purpose design.[28] For a user who does not need the added flexibility of the SYMAP program, the GRID program can provide substantial savings. The GRID program presently operates on an IBM 360/40.



II.1.b

GRID

For nominally scaled data, the map symbols are to be read as discrete types.

For quantitatively scaled data, the symbols are typically to be read as follows:

. = 0
, = 1
, = 2
+ = 3
X = 4
O = 5
Θ = 6
Θ = 7
▣ = 8
■ = 9

Example 3-10
Single Character Grid Cells (GRID)

Attractiveness for
Regional Recreation

Example 3-11
Multiple Character Grid Cells (GRID)

REPORTED BREAKING AND ENTERING
IN THE BOSTON SUBURBAN AREA (1969)
EXPRESSED PER THOUSAND POPULATION (1970)
BY CITY AND TOWN

9.33 - 20.46

7.61 - 9.33

6.06 - 7.61

4.15 - 6.06

2.10 - 4.15

Example 3-12
Conformant Map (CALFORM)

The Calform program[29] uses a pen plotter to produce maps similar to the SYMAP conformant type map, but with much greater precision and aesthetic quality (See Example 3-12). Because of the plotter's ability to move its pen in very small increments, finely detailed maps with subtle line spacing may be produced. The actual running time on a pen plotter may require several hours and as a result such maps are best prepared in limited numbers for use in final reports. This is in contrast to the SYMAP program which is better suited for volume production of maps for study. In this respect the two programs are quite complementary for the purpose of producing conformant maps. The Calform program operates on an IBM 360/40 and also requires the use of a pen plotting device.

3.3. AIR POLLUTION CONSIDERATIONS

3.3.1. Types of Maps

Most air pollution data is quantitative and assumed to be continuously variable in their geographic distribution. This includes surfaces of air quality from measured or modeled points. The handling of emissions data, however, is not so clear: area wide emissions should be represented as a continuous surface, but line or point source emissions should not.[30]

Emission densities for area sources have generally been shown by the zones--usually grid cells--for which the data have been gathered. Point and line sources would then be shown separately or else lumped together into the zones with the area sources. This latter approach is often misleading and certainly not an advisable one to use. Cells which

have major point or line sources within them appear to have much greater emission densities than adjacent cells that do not have such sources. The most meaningful display procedure would use a surface representation for area and probably line source emission densities, but would show point sources either on a separate map or as distinctive legends on the density map.

The classification of data values for mapping may take into account the general ranges of values useful to a control agency. On the other hand, particular values of interest--such as the air quality standards--may be singled out as the contour or isoline intervals. A graphic symbol or tone of gray would be assigned to each one of the levels or classes of values. When the values are assigned to classes based on an air quality standard one might have only two levels and therefore only two symbols. One level would show areas above the standard while the other area would show areas below the standard.

Contour or isoline maps provide the best representation for the majority of air pollution and related data. The conformant or zone map can be useful for qualitative data, such as opinion surveys, and for emissions data that has both been gathered by and is to be displayed by zones. The incidence or dot map would have more limited applicability but could be useful for displaying point sources or code enforcement information.

Symbol maps may be useful for displaying variable information <u>at a point,</u> such as the different quantities of fuel use or emissions from power plants. At the location of each power plant a circle would be shown, with size and/or

tone related to the quantity of fuel use or emissions. Clearly, as circles begin to overlap, this technique becomes less useful; one must be careful of the application chosen.

Three-dimensional representations of the contour or conformant type surfaces obviously have great applicability for purely visual display to government or civic groups. For analytical purposes they are only useful for a general understanding of the variability in the surface and for determining what classification of data values would be useful for the contour or conformant maps.

Four-dimensional maps can be very useful for information that varies in the Z-values direction. It would be particularly applicable to modeled or measured air quality if values were determined at several altitudes above the ground as well as at ground level. We could thus produce contour maps horizontally at the appropriate altitudes as well as vertically to show variations in the surface in that direction.[31]

Color is used with computer mapping mainly as an end product, particularly when two variables are to be displayed on the same surface. For analytical purposes during a study, these values would probably be displayed on separate maps. But, under certain circumstances, the different variables would be shown shaded in several colors on one map for the final presentation of the information.

Non-geographic mapping has very definite applicability to air quality data, particularly to data sets which show variability over time. The x and y axes might show recording stations for a region versus months of the year, respectively, with the air quality levels shown as the vertical or z-value.

We could thus see the general variation in the data for our entire sampling network without spatially locating each of the stations on a map. Upon examining this three-dimensional histogram, we might feel that a particular station was consistently reading too high or too low and check the values of this station before mapping.

3.3.2. Computer Programs

Only three computer mapping programs have been used in our air pollution studies: the SYMAP, SYMVU and GRID programs.[32] Plotter programs such as SYMVU are useful mainly for display. The preciseness of plotter output does not make up for the delays in plotting time or the relative scarcity of pen plotters.

For the majority of air pollution mapping applications the rapid production of maps on the line printer (with commensurate lack of detail) is more appropriate than plotter output. However, whenever presentation or display is the primary consideration, plotter programs and plotter output should be seriously considered. As more and more data becomes available as part of an on-line system, such as a telemetered air quality monitoring system, CRT plotters will have greater applicability in this field. Maps of present air quality and projected air quality as determined by a diffusion model can be shown quickly on a CRT screen, together with any appropriate statistical data. Emergency control decisions can then be made with a better understanding of the situation. This is particularly true for an agency which may institute different control measures for different parts of the region.

Printer programs such as SYMAP and GRID undoubtedly have the greatest applicability to air pollution data. The SYMAP program can be used to interpolate a surface from non-uniformally spaced data points. This flexibility is very useful for analytical purposes when one is concerned with air quality or emissions data. Once a base map has been made for an area, a data bank can be repeatedly accessed and a large series of maps run. The SYMAP program has proven to be very useful for producing series of air quality maps over time.

The GRID program, on the other hand, displays data collected on the basis of geographic grid cells. It is, therefore, useful for comparing maps or surfaces on a cell by cell basis. One might wish to compare air quality data gathered from a sampling network with population or related data gathered by census tracts. Such a comparison can show the areas where the air quality exceeds the standard and a large number of susceptible people (such as those presently ill with respiratory diseases) live.

4. DATA FOR GRAPHIC DISPLAY

In this section several types of data which might be used with a computer mapping system are discussed, as well as the methods by which such data could be stored and retrieved. Some of the techniques are applicable primarily to large data systems and may be feasible only in metropolitan areas.

4.1. TYPES OF DATA

Geographic data can be separated into three major categories, in terms of the coordinates required to identify location: point, line and area. An urban area may be shown, for example, as a point on a small scale map or as a bounded area on a large scale map. By increasing the scale of the map, some objects such as land parcels may still be shown as points. Each type of data requires its own method of processing and mapping.

4.1.1. Point or Node Data

Point or node data can be identified by a single coordinate pair which defines the map position of the phenomenon. For computer mapping purposes, data collected by the U.S. Census Bureau and by local government according to street address--combined with street address conversion and matching techniques--eventually will make point data the most important information in a computer mapping system. Presently this point data is rarely in machine-readable form.

4.1.2. Line or Segment Data

Line or segment data require two pairs of co-ordinates to describe the end points of the line.[33] The line may describe data of either a dynamic or a static nature. For example, one may represent the magnitude and direction of flows between two points, such as traffic volumes. One may also measure a purely descriptive statistic between two points such as the number of abutting houses recorded for a given street segment.

4.1.3. Area or Zone Data

Area data require that at least three points be known to describe a given zone.[34] Area data are usually a summation or ratio of some phenomena occuring within the zone. When the zones vary in size--as do city boundaries and census tracts--it is usually desirable to report the data as a ratio for the zone: percent, per capita, per square mile. The values for each zone can then be compared on the same basis, regardless of the different sizes of the zones.

4.1.4. Incompatible Data Zones

When working with area data for two or more subjects, one often finds that the data have been collected by different zones--such as school data by school districts and health data by health districts. The most popular solution to this problem is to place a relatively fine grid over the study area. Then each school district and each health district would appear as subdivided into common units--the grid cells. The datum value--which must be a ratio, as discussed in the previous section--is

then uniformly distributed throughout any given zone (school or health district), and this value is assigned to every grid cell which falls entirely or predominately within the zone.

A more refined technique may be used with quantitative data. A separate contour map may be prepared for each subject on the basis of its own data base. The common grid can then be placed over each contour map to derive the values for each grid cell. This technique eliminates the manual assignment of the original data zones to grid cells; the computer creates a surface from the data and determines the values for each grid cell. By generating a continuous surface from the data, the necessity of making the assumption that a data value should be uniformly distributed throughout a zone is eliminated.

The entire problem of incompatible data zones can be eliminated if data for each subject can be obtained initially for small areas. Then, the data can be aggregated into a common set of larger zones.[35]

4.2. DATA STORAGE

Much of the data used by a group such as an air pollution control agency will not have been collected by that agency. Data will be sought from as wide a variety of sources as possible. In every case, however, it must be possible to transform the coordinate locations of the data into a standard coordinate system, generally, latitude and longitude.

The three types of data differ in their storage needs. For all types, however, it is important to consider the efficiency of:

(i) adding new data to the file, as it becomes available;

(ii) retrieving data for mapping; and,

(iii) updating data in the future.

Storage media are not critical since the methods presented are equally applicable to card, tape or disk storage.

In the following sections it is assumed that the latitude and longitude coordinates, some quantitative or qualitative data, and an identifying number or name for the data recording unit are all available to the user.

4.2.1. For Point Data

There are two alternative methods for the storage of point data. In the first, and most commonly used, method the data point coordinates and associated variables are treated as a single unit or record.

> For example, one card would be used for each data point, and several data variables would be included on the card.

If a predefined partitioning or format is used for all records, the coordinates and associated variables for each data point can be retrieved for mapping purposes. As additional data is made available, it can be added to the end of each record by assigning a position to the new variable and modifying the record format. Updating is accomplished by replacing each data item with its new value.[36]

In the second method all the observations of a single variable are stored as a separate "package" with the variables

for each data point appearing in the same position in all of the packages.

> For example, one card would be used for each variable, and the values for several data points would be included on the card.

To retrieve data, only the packages for which data is desired are used, together with the corresponding coordinates. To add new data or to update existing data, a new package is prepared for the appropriate variable.

4.2.2. For Line Data

There is relatively little line data available to an air pollution agency at present. The notable exception is transportation data. Such data would be stored by a method similar to point data; however, there would be as many coordinate pairs as are required to define the line or curve, by line segments.

4.2.3. For Area Data

Area data has traditionally been represented by methods used for the storage of point data--a single pair of coordinates typically specifies the estimated centroid of a given area. The use of the centroid can be justified for statistical and analytical purposes, but it eliminates the possibility of reproducing the areas or zones from which the data was summarized and extracted. The storage of area data should reflect the possible need for mapping and analysis by the original zones; therefore, the coordinates of the vertices bounding each aerial unit should be retained.

Any of four methods can be used to record the locational character of each variable. The first two were developed for the storage of topographic information.[37] Either the value or intensity of the variable at uniform grid point locations, or, conversely, the x-y coordinates along selected isolines are recorded. Elegant methods of storage and retrieval of data for continuous surfaces have been developed to assist in the mapping of large amounts of military data; they seem applicable only to large scale, heavily endowed systems at present.

A third method involves the recording of the boundary coordinates for areas which describe the 'continuous extent' of a variable.[38] For example, the extent of each continuous soil type, land use class, or zoning category would be recorded, without reference to any pre-established data zone. It is then possible to compare the areas which were recorded for several different variables to determine the extent of their spatial correlation. This "map models" approach[39] is particularly well suited for recording qualitative data which is not area-specific, but rather can be imagined to "flow" over the landscape, as with soil types. In comparison to other methods, the number of records which are needed for any one variable is greatly reduced.[40]

The fourth method involves storing the coordinates of the boundary vertices for each data collection zone. It is well suited for quantitative data obtained from specified data zones. As with line data, a variable number of coordinate pairs would be required to describe each unique zone. Additional information, such as the centroid coordinates, would be stored either as a single unit or as separate packages and handled in a form similar to point data.

4.2.4. Conclusions

Point, line, and area data should be stored with one record per location with the variables related to the location on the same record. Since the area data of greatest value to air pollution control agencies is quantitative in nature and has been collected for pre-established data zones--such as area emission grid cells--the areas should be described in terms of the boundary vertices of the zone, or by the centroids if the zone sizes and shapes are uniform.

When all data for a given location cannot be stored on the same record without the records becoming extremely large and difficult to work with, several files of such records should be created. The contents of any one file should reflect the nature of the subject matter being recorded and the manner in which it will be used with other data.

4.3. AIR POLLUTION CONSIDERATIONS

4.3.1. Types of Data

As with any geographically specified information for a region air pollution data can be separated into three major catagories. These categories are in terms of the coordinates required to identify location: point, line, and area. Point data includes most measured and modeled air quality information and point source emissions. Line data is confined to mobile or motor vehicle source emissions, while area data includes emissions from area-wide sources.[41]

The question of incompatible data zones is very important in air pollution studies: air quality information is

generally gathered by sampling station or emission zone while socio-economic or land use data is gathered by census tract or land use zones. To compare the two types of data it is necessary to create a contour map for each subject on the basis of its own data base. The GRID program can then be used to compare the two surfaces cell by cell.[42] By generating a continuous surface from the data it is no longer necessary to assume that a data value should be uniformally distributed throughout a zone. This is particularly relevant for area wide emission data which has been mapped by zones with discontinuities at the zone boundaries. With a contoured surface there are no discontinuities at the zone boundaries.

4.3.2. Data Storage

The storage requirements of air pollution data should include provisions for a standard coordinate system and for the conversion of one coordinate system to another. Too frequently one set of air pollution data will be gathered by one coordinate system such as State Plane Coordinate or Universal Transverse Mercator while another set of air pollution or related data will be gathered by a different system.[43] It therefore becomes impossible to spatially compare the information. Common data specifications will become more widespread with national data banks and telemetered systems. This should greatly increase the compatibility of coordinate systems.

Storing on one record data point location and the values for all variables associated with the data point is most desirable for mapping purposes. However, air pollution data is rarely in this form. Generally all values for a variable--regardless of data point location--will be stored together as one package.

If the packages for different variables are to be used as a data bank for mapping--together with the appropriate data points--all values for a variable must be stored together in the same order in each package. This often necessitates data handling programs to transform the data from the order it has been stored in to the order necessary for computer mapping.

In many cases the data has been gathered, manipulated and stored without any regard for spatial location. In such cases one must start over again and determine the spatial locations for each value, store these in a separate package, and make sure that the order of the values for each variable is the same as the order of the data points package.

When one is concerned with area data such as area wide emissions from zones, it is important to store the coordinates of the boundaries of the zone and not just the centroid of the zone. If only this one location is specified it is impossible to use the data at a later time for zone mapping.

In certain cases the storage medium of the data is important. This is generally because of the lack of uniform data storage and handling specifications for air pollution data. The storage practices used from agency to agency make it quite difficult to work with data on tape.[44] It is therefore suggested that data be initially worked with on cards until such time as standard formats are used for all types of air quality data. This data can be placed on tape for dead storage as long as clearly understood and well documented formats are used. The data can be stored temporarily on disk when quick and inexpensive access by data handling programs is a factor.

4.3.3. Data Manipulation

Two types of data manipulation programs were developed in the course of the air pollution studies. Only the first type involves data point locations.

For several of the study areas it was necessary to aggregate data cells or data zones so as to have more reasonable numbers of data points for mapping. In the case of emissions data for the Connecticut Study,[45] previously defined grid cells were simply aggregated into larger cells. In the Puget Sound Study,[46] the emission zones were defined by the program and then the motor vehicle emissions placed in the appropriate zones.

The second type of data handling and manipulating program involved the time-averaging of data--the manipulation of the values to be mapped when the data points to be used for mapping were already known. This latter set of programs is in no way unique to computer mapping; it is necessary, however, for any statistical manipulation of air quality and related data.

(11) Unless the line happens to be horizontal, vertical, or at a 45° angle, the sequence of characters assumes a step-like appearance.

(12) usually 1/10" by 1/8" in size.

(13) at least with IBM 1403 printers.

(14) Plotters which draw with pens on paper come in two basic types, those that operate on a flat "bed" by moving one or more pens over the stationary paper and those that operate on a roll of paper by moving a pen from left to right while the paper moves up and down on a drum. The first type allows the use of preprinted paper (such as a base map) and fairly large sheets - from 30" by 40" to 5' by 24'. While the second type is limited to the width of the paper spool--12" to 30"--its almost unlimited length is useful for linear displays such as strip charts.

(15) The image is transferred to film or photosensitive paper which must then be developed to obtain 'hard copy'.

(16) Most CRT devices are limited to a screen size of 10" by 10" with 10,400 possible raster cells per square inch.

(17) Assuming a minimum character size of 0.1" by 0.1", each symbol requires at least 100 raster units. This would result in a matrix of 102 by 102 for the effective number of character display positions on a 10" by 10" screen.

(18) Ten to twelve such tones appear to be maximum, with five to seven being optimum for most purposes. It is desirable to use symbols (e.g., letters) which provide a visual continuum of tone when clustered, but can be differentiated from the symbols of other value levels upon close inspection.

(19) Robert Stevens, using data gathered at four crossings west of the Galapagos Islands, produced the three-dimensional model illustrated here, showing the east-west velocity of the Cromwell Current within a section of the Pacific Ocean. The data was given in four vertical planes located at 93, 96, 118, and 140 degrees west longitude, extending from 2 degrees south latitude, to a depth of 200 meters. Each plane represented a vertical "slice" of ocean water. By placing a regular twenty-five point grid on each of the given planes (which were of equal size), it could be assumed that each point was the intersection of three mutually perpendicular (orthogonal) planes, and thus an additional set of points and values was gained for five vertical latitudinal

planes and for five horizontal subsurface planes. By estimating velocities from the contours given in the original data, a series of fourteen SYMAPs was prepared for mounting in three-dimensions on an assembly of small wood blocks. See: Report by Robert Stevens on file at the Laboratory for Computer Graphics and Spatial Analysis.

(20) The maps should use the same electives so that size, symbolism, and value range intervals are held constant.

(21) Population distribution in the United States in 1960 is shown as a three-dimensional histogram drawn on an incremental pen plotter. Based on county statistics provided by Dr. William Warntz, population data are aggregated into one degree quadrilaterals of latitude and longitude. Each quadrilateral is raised to a height proportional to its population. The program removes hidden lines by plotting from front to back and drawing only those line segments whose y-axis value is higher than any line previously plotted at a corresponding x-axis location. A non-linear vertical scale of population is used for better visibility of lower value zones. The program was developed by Frank Rens at the Laboratory.

(22) Plotter programs, have been prepared by the Navy's Atlantic Intelligence Center in Norfolk, Virginia, the Department of Geography at the University of Michigan, the Puget Sound Regional Transportation Study, the Systems Development Corporation (as part of the SPAN computer system), and others.

(23) Horwood's and the Map 01 programs are written for an IBM 1401 computer; 1401 computers are relatively inexpensive and have been widely used by local governments.

(24) As output on the line printer each 1/2" square grid cell contains 20 possible print positions--5 across and 4 down.

(25) The concept, overall design, and mathematical model for the SYMAP program were developed in the autumn of 1963 by Howard T. Fisher, working at Northwestern Technological Institute. Programming was carried out by Mrs. O. G. Benson of the Northwestern University Computing Center. Since that time many others have contributed ideas and cooperated in bringing the program to its present state. Major improvements in the program were made by two Harvard students, Robert A. Russell and Donald S. Shepard,

who was responsible for the algorithms on spatial interpolation. The program is available from the Laboratory.

(26) The SYMVU program was written at the Laboratory for Computer Graphics and Spatial Analysis in 1967 and 1968 by Frank J. Rens under the direction of Howard T. Fisher. Since then it has undergone major changes and improvements.

(27) The GRID program was written at the Laboratory for Computer Graphics and Spatial Analysis in 1968 by David Sinton and Carl Steinitz. The program is available from the Laboratory.

(28) The basic difference is as follows: GRID requires that a value be supplied for every cell on a uniform grid. SYMAP interpolates a surface from non-uniformally spaced data points and thereby derives a value for each cell on a grid. The GRID program is more efficient because it does not use the relatively expensive interpolation procedure.

(29) The CALFORM program was written at the Laboratory for Computer Graphics and Spatial Analysis in 1970 by Robert S. Cartwright, Jr.

(30) We mapped point and area emissions data for the Houston region (using the NAPCA inventory developed to define the intrastate Air Quality Control Region) using both contour and conformant mapping and specifying point sources either as legends or as conformant zones. No conclusions were reached as to the best presentation method.

(31) Modeled air quality data at various altitudes is available from several studies, notably one undertaken for point source power plant emissions by the Tennessee Valley Authority.

(32) The GRID program was used only for the St. Louis and Kansas City Studies (see Volume Two, Reports A, B and D). The SYMVU program was used in these studies and several others as well, including the Southern New England Study (Report G).

(33) The coordinates for line data should not be confused with boundary coordinates which describe the perimeter of an area or zone of interest.

(34) The actual number of points is equal to the number of straight line segments required to approximate the boundary of the zone. The coordinates of the end points for each line segment thus describe the location and extent of a data zone.

(35) Of greatest value is data which is not aggregated at all. If data is available just as it was recorded originally--for example, by street address--it can be aggregated into any larger area unit. Clearly, this gives complete flexibility. This was the principle followed by the U.S. Census Bureau in its New Haven Study. It is possible to assign data to a block face and then aggregate block faces into any area desired for the reporting of census data.

(36) In both operations, a complete search of the data bank is necessary and can become very time consuming for a larger data bank.

(37) This implies that the data is taken from a continuous surface.

(38) An area is defined such that the value for a particular variable is constant throughout--e.g., land use is all single-family residential. This area is approximated by a polygon. The vertices of the line segments forming the polygon are then coded as the boundary coordinates.

(39) The approach has been extensively developed by the Portland Metropolitan Planning Commission and the Bureau of Municipal Research and Service at the University of Oregon and is called Map Models. Unfortunately, a separate map model must be prepared for each variable to be mapped. As a compromise, a map model is prepared for a limited number of phenomena such as land use, and additional map models are made of data zones such as blocks or census tracts for which a variety of data is available.

(40) For a large number of variables, a map model file would eventually exceed a parcel file in size.

(41) We did not work with any line source data in our studies; in the case of the Puget Sound Study (Volume Two, Report H) motor vehicle emissions were aggregated into grid cells and expressed in terms of area emission densities. Likewise we did not work with any point source data directly. Point source emissions were either included in area-wide emission densities or excluded from the data mapped. All air quality and related data were mapped as surfaces interpolated from data points; no data were mapped strictly by zones.

(42) See the comparison mapping in the Kansas City Study (Volume Two, Report D)

(43) For example the air quality data used in the Southern New England Study (Volume Two, Report G) had been gathered by a different coordinate system for each state (each state has its own State Plane Coordinate System). The population and income data was available by still another coordinate system.

(44) The Air Quality data used in the St. Louis Region Study was provided 'packed' on a tape incompatible with our computer equipment and programs. A great deal of money and time was expended in deciphering this data tape; all information on cards is "compatible".

(45) See Volume Two, Report F.

(46) See Volume Two, Report H.